UNIVERSITÀ DEGLI STUDI DI MILANO

SCUOLA DI DOTTORATO
Scienze Molecolari e Biotecnologie Agrarie, Alimentari ed Ambientali

Dipartimento di Endocrinologia, Fisiopatologia e Biologia Applicata
Facoltà di Farmacia

CORSO DI DOTTORATO
Biotecnologia degli Alimenti/ciclo XXII

TESI DI DOTTORATO DI RICERCA:

Lupin proteins as food ingredients: development and optimization of label-free methods for the quantitative analysis in mass spectrometry

sigla dei settori scientifici disciplinari
CHIM 10/AGR 15

NOME DEL DOTTORANDO
Donatella Resta
Matr R07160

NOME E COGNOME DEL TUTOR
Anna Arnoldi

NOME E COGNOME DEL COORDINATORE DEL DOTTORATO
Luciano Piergiovanni

A.A. 2009-2010

INDEX

Chapter 0	**PREFACE**	1
0.1	**References**	3
Chapter 1	**STATE OF THE ART**	4
1.1	**Lupin**	5
1.2	**Lupin proteins**	6
1.2.1	Globulins	8
1.2.2	β-conglutin or vicilin like-protein	9
1.2.3	γ-conglutin	11
1.2.4	α-conglutin or legumin-like protein	13
1.2.5	δ-conglutin	13
1.3	**Mass Spectrometry**	15
1.4	**HPLC-Chip**	17
1.5	**Electrospray ionization**	18
1.6	**Electrospray (ESI) and Nano-Electrospray (nano-ESI)**	21
1.7	**Mass analyzer**	22
1.7.1	Quadrupole ion trap	22
1.7.2	Linear two dimensional ion trap (LTQ)	24
1.7.3	LTQ-Orbitrap	25
1.8	**Proteomics**	26
1.9	**Shotgun Proteomics**	27
1.10	**Mass Spectrometry instrumentation for Shotgun Proteomics**	29
1.11	**Fragmentation techniques**	29
1.12	**Quantitative proteomics**	34
1.12.1	Relative quantitative proteomics	34
1.12.1.1	Stable isotope techniques	34

1.12.1.2	Stable isotope label free techniques	36
1.12.2	Absolute quantitative proteomics	39
1.13	**Quantitative proteomics and nutraceutics**	41
1.14	**References**	43
Chapter 2	**Study 1:A LABEL-FREE INTERNAL STANDARD METHOD FOR THE DIFFERENTIAL ANALYSIS OF BIOACTIVE LUPIN PROTEINS USING NANO HPLC-CHIP COUPLED WITH ION TRAP MASS SPECTROMETRY.**	51
2.1	**Aim of study 1**	51
2.2	**Flow scheme of study 1**	52
2.3	**Materials and Methods**	53
2.4	**Results**	56
2.4.1	Quantitative analysis of the storage proteins	56
2.4.2	Label-free differential analysis of the seed storage proteins	64
2.4.2.1	Quantitative analysis of the internal standard BSA	67
2.4.2.2	Quantitative analysis of γ-conglutin	69
2.4.2.3	Proof of the concept	75
2.4.2.4	Quantitative analysis of the vicilins	75
2.4.2.5	Proof of the concept	80
2.5	**Conclusion of study 1**	81
2.6	**References**	83
Chapter 3	**Study 2: LABEL-FREE ABSOLUTE QUANTIFICATION OF THE γ-CONGLUTIN IN LUPIN FLOUR: DEVELOPMENT AND OPTIMIZATION OF A NANO-HPLC-CHIP MULTIPLE REACTION MONITORING (MRM) METHOD BASED ON PROTEOTYPIC PEPTIDES.**	85
3.1	**Aim of study 2**	85
3.2	**Flow scheme of study 1**	87

3.3	**Materials and Methods**	89
3.3.1	γ-conglutin purification	89
3.3.2	HPLC-Chip–MS/MS analysis of the purified γ-conglutin	91
3.4	**Results**	92
3.4.1	Proteotypic peptides and HPLC-Chip-MRM	92
3.4.2	Development: standard addition strategy	94
3.4.3	Optimization: exogenous internal standard BSA	104
3.5	**Conclusion of study 2**	118
3.6	**References**	119
Chapter 4	**CONCLUSION**	120
4.1	**References**	123
Appendix 1	**Abstract**	124
Appendix 2	**Riassunto**	125
Appendix 3	**Papers**	126
Appendix 4	**Abstract of posters**	127
4.1	**Poster 1**	128
4.2	**Poster 2**	129
4.3	**Poster 3**	130
4.4	**Poster 4**	131

O. PREFACE

The primary function of dietary proteins is to adequately supply the body with essential amino acids and organic nitrogen, but some protein components could also be a potent source of biologically active peptides. Such regulatory peptides can be released by the gastrointestinal enzymatic proteolysis and may act as potential physiological modulators of metabolism. The quantification of bioactive or allergenic lupin proteins represents the main goal of my studies.

Mass-spectrometry based shotgun proteomics is the tool used to develop innovative analytical methods able to simultaneously trace and quantify target proteins. In this technique, the proteins are digested with a proteolytic enzyme to generate shorter peptides that are more easily analyzed by mass spectrometry. Shotgun proteomics relies on the separation of these peptides by reversed phase chromatography (RP-HPLC) directly coupled with tandem mass spectrometry analysis (MS/MS). Tandem mass spectrometry enables to infer the correct amino acid sequence of each peptide starting from the corresponding fragmentation (MS/MS) spectrum. Consequently, shotgun proteomics enables an unquestionable identification of the protein by sequencing its primary structure and, moreover, to trace proteins in complex mixtures without a previous isolation and purification.

For the relative quantification of target protein in mixtures, two main approaches are reported in literature: stable isotope labeling techniques (SIL) and stable isotope label free techniques (SIF). SIL techniques, in spite of their potency, have some limitations, since most of them are able to compare different protein expressions of a limited number of samples and, in addition, all require a proper chemical reaction to be performed before the proteomics analysis. It is very often unclear how the efficiency of these reactions and the capturing techniques used to enrich the samples of labeled peptides, may affect the quantification of the different proteins. Moreover, these methods can be applied only to those proteins containing the amino acid modified by the tagging reagent and the labeled reagents are always very expensive.

Although SIL techniques remain the core technology, increasing efforts have been directed to label-free methods which are promising alternatives to SIL techniques, especially in the field of food quality, due to their relatively simple workflow and to their capability to trace the differential expression of a target protein, potentially in an unlimited number of sample. In principle, the intensity of a given peptide chromatographic peak depends linearly on its concentration. This is not always true when analyzing real complex samples, where two peptides in equimolar amounts may show different intensities because of matrix effects due to ion suppression in mass spectrometer source. The introduction of an exogenous internal standard protein at a constant level in the protein mixtures subjected to enzymatic digestion seems to be an interesting solution both for evaluating matrix effect and for normalizing quantitative parameter of peptides and proteins, in addition to the use of a very sensitive and reproducible chromatographic system [Chambers et al., 2007].

In the first study an internal standard based SIF method was developed with the aim to trace the relative abundances of the main bioactive lupin proteins, vicilins and gamma-conglutin, in different cultivars of *Lupinus albus*.

The aim of the second study was the development of a reliable quantitative HPLC-Chip-MS/MS method for the absolute quantification of gamma-conglutin. Gamma-conglutin is a mature lupin protein composed by a heavy and a light chain linked by disulfide bonds [Blagrove et al., 1980; Restani et al., 1981]. This is a bioactive protein with possible

hypoglycemic activity [Magni et al., 2004]. Moreover gamma-conglutin is relevant also for another reason; a few literature data indicate, in fact, that some individuals are allergic to lupin proteins. This evidence prompted the European Commission to include this seed in the list of food allergens whose declaration on food label is compulsory. Some literature investigations have indicated that gamma-conglutin may be one of the major lupin allergen. This hypothesis is supported by some peculiar physical and chemical properties, such as the thermal stability and the resistance to proteolysis, that are features frequently shared by most food allergens.

Nowadays, the quantification of food allergens is generally based on immunoassays. These methodologies have some limitations, such as cross-reactivity with other food proteins and false-positive results [Marthy et al., 1998]. Mass spectrometry-based shotgun proteomics enables the simultaneous identification and quantification of the protein through the selection of target peptides.

A relative approach does not enable the absolute quantification of the target protein in the sample and, consequently, lacks to give a real evaluation of the potential bioactivity of the food. In order to achieve a real absolute quantification of the lupin gamma-conglutin, a very selective method was develop and applied to Total Protein Extracts (TPE) obtained starting from lupin seed flour.

The method is based on Multiple Reaction Monitoring (MRM) mass spectrometry and uses a standard addition of a well characterized exogenous protein for the absolute quantification.

The development of the MRM method for the quantification of a protein requires some preliminary phases including the evaluation of the efficiency of the protein digestion, the selection of the target peptides both of the bioactive protein, gamma-conglutin, and of the internal standard protein, the preparation of a calibration curve of the target protein in absence and in presence of matrix (i.e. respectively “standard calibration curve” and “matrix calibration curve”) in which the target protein has to be quantified, and the evaluation of the analytical parameters of the method (linearity, limit of detection and quantification, precision, accuracy, reproducibility).

0.1 References

Blagrove RJ et al., 1980, Physiochemical studies on conglutin γ a storage globulin from seeds of *Lupinus angustifolius*. Aust J Plant Physiol 7:1-13.

Chambers E et al., 2007, Systematic and comprehensive strategy for reducing matrix effects in LC/MS/MS analyses. J Chromatog. B 22-34.

Magni C et al., 2004, Conglutin gamma, a lupin seed protein, binds insulin in vitro and reduces plasma glucose levels of hyperglycemic rats. J Nutr Biochem 15:646-650.

Marthy JN et al., 1998, Tacrolimus metabolite cross-reactivity in different tacrolimus assay. Clin Biochem 31:613-617.

Restani P et al., 1981, Subunit composition of the seed globulin from seeds of *L. albus*, Phytochemistry 20:2077-2083.

1 STATE OF THE ART

A food can be regarded as functional if it is satisfactorily demonstrated to affect beneficially one or more target functions in the body beyond adequate nutritional effects, in a way that is relevant to either an improved state of health and well-being and /or reduction of risk of disease. Innovative nutritional strategies to reduce the main risk factors have been developed including either dietary changes or consumption of specifically targeted functional foods and dietary supplements. These nutraceutical products may also provide an alternative to lipid lowering, antihypertensive, and antidiabetic drugs. Functional foods and beverages have the appearance of normal foods, but contain specific components whose activity on at least one measurable risk factor has been scientifically demonstrated. Dietary supplements, having formulations similar to drugs, allow the delivery of a bioactive ingredient in dosages that exceed those obtainable from food products. Among bioactive components, at present dietary proteins from both vegetable and animal sources are of high interest, because of their specific effects on cholesterolemia and blood pressure.

The US functional foods market will grow by up to 20% yearly, according to a report from Pricewaterhouse Coopers that puts the market at $ 27bn in 2007. Another research conducted by the International Food Information Council (IFIC) found that of those Americans trying to improve their diets, 79% are changing the type of foods they eat, 69% are changing the amount of foods consumed, and 19% are changing their use of dietary supplements. In that study, the "top" functional foods named by consumers are: fruits and vegetables, fish/seafood, dairy (including milk and yoghurt), meat and poultry, herbs/spices, fiber, tea and green tea, nuts, whole grains, cereal, oat bran, and vitamins/supplements.

Among vegetables, leguminous plants are probably the best protein sources because of their high protein content and nutritional value. Plant proteins have an important role in the diet since they have been shown to induce a significant reduction of cholesterolemia both in experimental animal models and hypercholesterolemic humans [Sirtori et al., 1998; Bakhit et al., 1994; Anderson et al., 1995].

The cholesterol-lowering effect, potentially leading to a reduced cardiovascular risk, was the basis for the U.S. Food and Drug Administration (US FDA) approval of the health claim concerning the role of soybean proteins in reducing the risk of coronary disease [FDA, 1999]. This claim affirms that the consumption of 25 g of soy protein per day is useful in the prevention of cardiovascular diseases; because of this, soy proteins have become a successful ingredient in the preparation of functional foods for the prevention of cardiovascular diseases. All these facts have stimulated researchers on other legume species in order to select additional sources of bio-functional ingredients.

The interest in lupin protein was born for several reasons. First of all for the need of vegetable protein sources alternative to soy: in fact, the consumption of soy is accomplished with the

intake of not negligible amount of isoflavones, for which the pharmacological and, overall, toxicological properties have not been clarified and are still controversial. In addition, unlike soy, lupin is not associated with genetically modified organisms (GMO) leading to a great compliance for the consumer, and, moreover, it is the only other legume with a protein content similar to that of soy.

1.1 LUPIN

In the last few years the lupin have attracted the attention of research for its nutraceutical properties. Lupin is an herbaceous annual plant belonging to the other of *Leguminous*, *Papillionaceae* family, *Genesteae* section, *Lupinus* gender.

The lupin seed is produced in pods which develop on the main stem of the lupin plant.

Three species are cultivated: white lupin (L*upinus albus*), yellow lupin *(Lupinus luteus*), narrow-leaf lupin (*Lupinus angustifolius*) and pearl lupin (L*upinus mutabilis*)

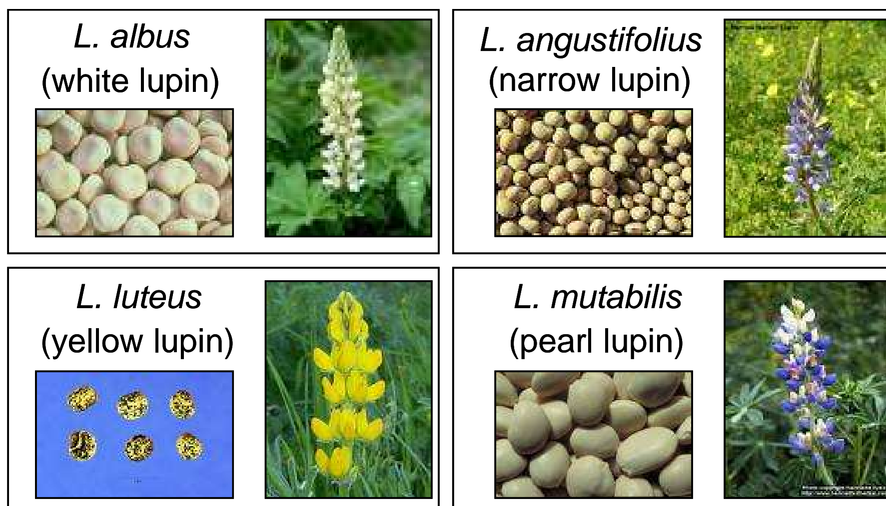


Figure 1.1: Four domestic species of lupin

Actually plant proteins are increasingly used as food ingredients because they improve nutritional profile, stabilise the texture and optimise recipe costs. Analyses of nutritional values

of sweet lupin have shown that the bioavailability of the constituents is comparable to those of processed soybeans [Blagrove et al., 1975]. In contrast to other leguminous plants (peas, soy, beans), lupins contain extremely low amounts of trypsin inhibitors, lectins, isoflavones, saponins and cyanogens.

1.2 LUPIN PROTEINS

The lupin seed is very rich in protein (34-43% of dry matter), has an important percentage of oil (5.4-10%) and contains very little starch (0.7-2.2%). The percentage of crude fibre is high (14-16.5 %). In addition the seeds of lupin species contain small amounts of a variety of other components, such as phytates, oligosaccharides and trypsin inhibitors. Traditionally these were known as antinutritional factors, but increasingly they are thought to be favourable bioactive factors because of their potentially beneficial effects in pharmacological, medical, cosmetic and food applications. The high protein content and relatively high oil content of lupin seed, that in case of white lupin is similar soybeans, makes the lupin crop interesting for human consumption as well as animal feed. Table 1.1 shows the different content of protein in different species of lupin seeds. The general amino acidic profile of lupin protein is comparable to that of soy protein; in fact, it is rich in leucine, lysine, and phenylalanine. However, similarly to the other leguminous, it is poor in sulphur amino acid such as cystein and methionine (Table 1.2).

Table 1.1: Protein composition in different lupin species.

	L. albus	***L. angustifolius***	***L. luteus***	***L. mutabilis***
Protein (%)	34.4-44.9	28.0-37.9	36.0-47.6	31.7-45.9

Seed proteins are almost totally localized in the embrional axis and in the cotyledons, while only about 2.5% are in the tegument [Hove et al., 1974].
The main protein fractions of lupin seeds are albumins (with catalytic activity) and globulins (storage proteins), both rich in glutammic acid, aspartic acid, arginin, and leucin. In addition, the absence of prolamines makes lupin an interesting source of proteins suitable as food ingredients dedicated to celiacs.

Table 1.2: Amino acid content (mg aminoacid/g protein) in different lupin species and in soybean.

Amino acid **mg aa/ g protein**	***L. albus***	***L. angustifolius***	***L. luteus***	***L. mutabilis***	***Soy***
Isoleucine	41.0	34.0	37.4	44.7	53.0
Leucine	71.0	58.2	71.4	74.3	91.0
Lysine	48.7	50.0	51.0	55.5	75.0
Methionine +Cysteine	14.7	9.0	19.0	14.0	30.0
Phenylalanine Tyrosine	75.7	71.2	57.0	92.7	95.0
Threonine	31.7	33.5	31.0	38.8	45.0
Tryptophan	58.2	70.0	58.0	10.5	15.0
Valine	37.7	38.0	35.0	38.3	56.0

Recently it has demonstrated that feeding a rat model of hypercholesterolemia with lupin proteins has beneficial effects on the total and LDL-cholesterol levels, in a way similar to soy proteins. Subsequently, this observation was strengthen by Wait et al. [Wait et al., 2005] who highlighted the great similarity between the primary sequence of the 7S lupin globulin (β-conglutin) and that of one main bioactive soybean component, i.e. the α' subunit of β-conglicinin.

The general amino acidic profile of lupin protein is comparable to that of soy proteins (table2); in fact, it is rich in leucine, lysine, and phenylalanine. However, similarly to the other leguminous, it is poor in sulphur amino acid such as cystein and methionine.

The main protein fractions of lupin seeds are albumins (catalytic activity) and globulins (storage proteins), both rich in glutammic acid, aspartic acid, arginin, and leucin. Some authors reported the existence of a fraction soluble in alkaline solution, smaller than the first two, known as glutelin. In addition, the absence of prolamines makes lupin an interesting source of proteins for gluten-free dietary products.

Albumins are soluble in weak acidic solution (pH 5.0-5.5), and represents about 10% of the total protein content. They compose a very heterogeneous system, and the most of these are

involved in enzymatic activity of the metabolism of seeds, although some others seem to be storage proteins.
In this thesis, only storage proteins will be considered

1.2.1 Globulins

Globulins are storage proteins, with no catalytic or structural properties, that undergo hydrolysis during the germination stage providing carbon and nitrogen to the plant.
They are essentially stored in protein corpuscles in the parenchymal cells of cotyledons .
Globulins represent about 87 % of the total protein content of lupin seeds. They are insoluble in water as well as in alkaline solutions [Duranti et al., 1997].
The first globulin classification allowed the identification of four fractions know as α-, β-, γ, δ-conglutin [Blagrove et al., 1980]. For the similarity with vicilin and legumin proteins of pea, β-conglutin is also known as vicilin-like protein, while α-conglutin is known as legumin-like protein.
In the past, lupin proteins were classified on the basis of their sedimentation coefficient:

- 7S globulin: β- conglutin or vicilin-like proteins, with a sedimentation coefficient between 7 and 8S.
- 11S globulin: α-conglutin or legumin-like proteins, with a sedimentation coefficient of 11S.
- 2S globulin: δ- conglutin with a sedimentation coefficient of 2S.

Generally, globulins are oligomeric proteins denaturated in the presence of denaturing agents such as urea or SDS, liberating the individual chains of the monomers. All the globulins derived from a unique common ancestor polypeptide, which undergoes proteolytic cleavage giving a complex mixture of polypeptides, which aggregate to form globulins. In addition to this, the immature proteins are subjected to many post-translational modifications. Due to these phenomena, globulins are an extremely heterogeneous class of proteins. Despite of this high heterogeneity, it was possible to identify legumins and vicilins in numerous classes of vegetable, i.e. legumes, cereals, etc.
This fact suggests that, probably, the gene codifying for globulins is an ancestor one, already existing before vegetables differentiate in species [Duranti et al., 1997]. Therefore, legumins and vicilins have in common the same ancestral origin [Borrato et al., 1987; Gibbs et al., 1989]. In fact, even if deep differences exist between the various classes of globulins, they have a highly conserved sequence and similar structural relationships. For this reason, globulins can be considered as protein markers for the different species of legumes [Duranti et al., 1997].

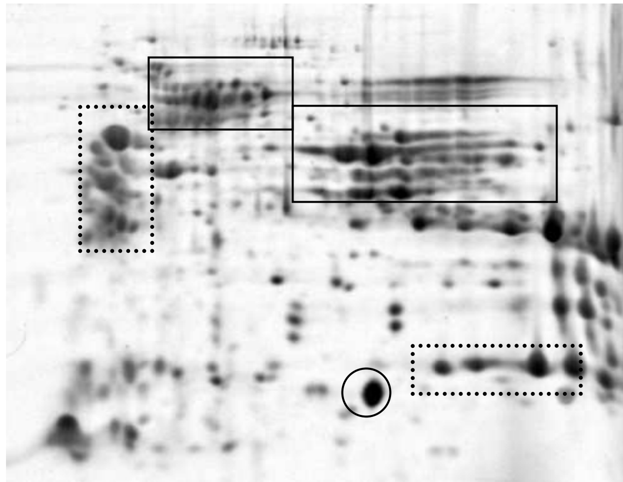


Figure 1.2: 2D electrophoresis of the total protein extract obtained by white lupin seed. Whereas γ-conglutin is a homogeneous protein, composed by a heavy and a light chain linked by disulfide bonds (circles), the vicilins(solid line rectangles) and the legumins (dashed line rectangles) have a multigenic origin and appear as complex mixtures of polypeptides with different molecular weights and pIs. The lupin protein sequences deposited in the database NCBInr were the following: for the class of vicilins, the β-conglutin precursor (NCBInr accession no. 46451223) and the vicilin-like protein (NCBInr accession no. 89994190); for γ-conglutin, the sequence NCBInr accession no. 11191819; for the class of legumins, the legumin-like protein (NCBInr accession no. 85361412); for α-conglutin, the δ-conglutin seed storage protein precursor (NCBInr accession no. 80221495).

1.2.2 β-conglutin or vicilin like-protein

Vicilin-like are generally trimeric proteins (MW 150-170 KDa). Each monomer has a relative MW between 40-75 KDa, and they are associated each other throughout weak bond, overall electrostatic force, between the side chain of the amino acidic residuals. Because of this, the

association-dissociation equilibrium of β-conglutin is affected by ionic force and the pH of the medium.
No cysteines are present in the primary structure of such a protein, so that no disulfide bond can be done between the subunits [Duranti et al., 1997]. Vicilins are glycoproteins mainly containing mannose units covalently linked to the different subunits.
During cotyledon development, β- conglutin is synthesized as a precursor with MW of 64 KDa; this one disappears during seed maturation forming the typical polypeptides, which will compose the single subunits in the mature proteins [Duranti et al., 1997].
Studies related with the catabolism of storage proteins in seed of *L. albus* have evidenced that α, β- and γ-conglutin undergo to proteolytic cleavage during germination and growth of the plant. More specifically, γ-conglutin has a slow proteolysis, α-conglutin has an intermediate degradation and the new peptides coexist with those of the dry seed, and finally β-conglutin is the most hydrolyzed protein [Ramos et al., 1997]. The proteolitic process of such a protein is so intense to generate a series of polypeptide . The accumulation of an intermediate peptide (MW 20KDa) originated from β-conglutin is of particular interest. In fact, this seem involved in the defensive system of the plant, havin antifungal activity and toxicity against insects [Ferreira et al., 2003].
The amino acid sequence of β-conglutin is reported in figure 1.3.

Beta-conglutin precursor (gi|46451223)
1 mgkmrvrfpt lvlvlgivfl mavsigiayg ekdvlksher peereqeewq prrqrpqsrr 61 eereqeqegg spsyprrqsg yerrqyhers eqreereqeq qqgspsysrr qrnpyhfnsq 121 rfqtlyknrn gkirvlerfd qrtnrlenlq nyrivefqsk pntlilpkhs dadyvlvvln 181 gratitivnp drrqaynley gdalripags tsyilnpddn qklrvvklai pinnpgyfyd 241 fypsstkdqq syfsgfsrnt leatfntrye eiqrillgne deqeyeeqrr gqeqshqdeg 301 vivrvsreqi qeltkyaqss sgkdkpsqsg pfnlrsnepi ysnkygnfye itpdrnpqvq 361 dldisltfte inegalllph ynskaifivv vgegngkyel vgirdqqrqq deqeeepeev 421 rrysarlseg difvipagyp isvnassnlr llgfginaye nqrnflagse dnvirqldre 481 vkeltfpgsa edierliknq qqsyfanalp qqqqsekeg rrgrrgpiss i
Vicilin-like protein (gi|89994190)
1 makmrvrlpm lilllgvvfl laasigiayg ekdftknppk ereeeehepr qqprprqqee 61 qerehrreek hdgepsrgrs qseesqeeeh errrehhrer eqeqqprpqr rqeeeeeeee 121 wqprrqrpqs rreereereq eqgsssgsqr gsgderrqhr errvhreere qeqdsrsdsr 181 rqrnpyhfss nrfqtyyrnr ngqirvlerf nqrtnrlenl qnyriiefqs kpntlilpkh 241 sdadfilvvl ngratitivn pdkrqvynle qgdalrlpag ttsyilnpdd nqnlrvakla 301 ipinnpgkly dfypsttkdq qsyfsgfskn tleatfntry eeiervllgd delqenekqr 361 rgqeqshqde gvivrvskkq iqelrkhaqs ssgegkpses gpfnlrsnkp iysnkfgnfy 421 eitpdinpqf qdlnisltft einelwltke keimn

Figure 1.3: Amino acid sequence of β-conglutin precursor and vicilin-like protein of *L. albus*, from NCBI database.

1.2.3 γ-conglutin

The γ-conglutin is a glycoprotein present in many species of Lupinus, and it constitutes about 6% of the total proteins in mature seeds of white lupin. It is an oligomeric protein, synthesized in the first stage of development of the seed starting from a precursor peptide (MW 47-51 KDa). It is a tetrameric protein, whose single monomers (MW 47 KDa) are composed by two chains, called heavy (MW 30 KDa) and light chain (MW 17 KDa), linked one another by disulfide bonds. The heavy chain is glycosylated [Lui et al., 2004] and shows a glycated residue formed by mannose and glucosamine, whereas the light chain is not glycosylated [Restani et al., 1981]. The unusual resistance to proteolysis of this protein, both *in vivo* during germination and in vitro with proteolytic enzyme (i.e. trypsin), was initially related with a general protective

effect of the polysaccharidic chains on the protein itself [Semino et al., 1985]; however, more recent studies have highlighted that those chains are not directly involved in the protection of native protein against proteolysis, but they increase the refolding rate of denatured protein [Duranti et al., 1995]. More probably, such a resistance is due to the presence of a great number of disulfide bonds, both inter- and intra- chain, between the 13 cysteines present in the primary structure of γ- conglutin, resulting in a super-coiled protein.
On the basis of a series of considerations, γ-conglutin has been considered a storage protein for a long time: it is abundant in seeds, where it is synthesized and accumulated in the protein corpuscles of parenchimatous cells, and in addition no biological activity seems to be related with such a protein.
However, γ-conglutin has some typical characteristics wandering it from globulins. In fact, it is characterized by a different amino acidic profile, being rich in lysine, threonin, tryptophan, and sulphur amino acids; it is particularly resistant to heating treatment; it is secreted during the germination of seed [Duranti et al., 1994; Duranti et al., 1995]; finally, studies of immunocytolocalization have shown that, differently from the other storage globulins, this protein is located in epidermal cells in mature seeds or intercellular spaces associated with lipidic globes during germination. Some years ago, it was observed a very high sequence homology between γ-conglutin and the Bg7S of soy. The latter protein has behaviour similar to the lupin protein, being secreted during the germination phase and after thermal treatments [Hirano et al., 1987]; furthermore, it is able to bind insulin and insulin-like growth factors, and it seems to have a tyrosin-kinase activity. For these reasons, the researches have hypothesized a proper catalytic activity for γ-conglutin, in particular a lectin-like one [Duranti et al., 1995].
The amino acid sequence of γ-conglutin is reported in figure 1.4.

Conglutin gamma (gi|11191819)

```
  1 marnmahilh ilvislsysf lfvssssqds qslyhnsqpt sskpnllvlp vqedastglh
 61 wanihkrtpl mqvpllldln gkhlwvtcsq hysssstyqap fchstqcsra nthqcftctd
121 stttrpgchn ntcgllssnp vtqesglgel aqdvlaihst hgsklgpmvk vpqflfscap
181 sflaqkglpn nvqgalglgq apislqnqlf shfglkrqfs vclsrystsn gailfgdind
241 pnnnnyihns ldvlhdlvyt pltiskqgey fiqvnairvn khlviptknp fispsstsyh
301 gsgeiggali ttthpytvls hsifevftqv fannmpkqaq vkavgpfglc ydsrkisgga
361 psvdlildkn davwrissen fmvqaqdgvs clgfvdggvh aragialgah hleenlvvfd
421 lersrvgfns nslksygktc snlfdlnnp
```

Figure 1.4: Amino acid sequence of γ-conglutin of *L. albus*, from NCBI database.

1.2.4 α-conglutin or legumin-like protein

Legumins represent about the 33% of the total protein content of white lupin seed; globulins 8S and 9S establish the 64 and 36% of this fraction, respectively. As highlighted before, legumins, more rich in sulphur and essential aminoacids, have a better nutritional value. They are oligomeric proteins formed by 4 kinds of monomers. Through non denaturing SDS-page, it has been possible to determine the MWs of those, which resulted 81,75,72, and 69 KDa.
After adding of denaturing agents, a similar procedure have evidence that each monomer was composed of two subunits, an acid and a basic chain linked each other by disulfide bonds [Duranti et al., 1997; Melo et al., 1994]. The α-conglutin is a glycoprotein, glycosylated in the acidic chain, with the glycated portion more represented in the 8S than in the 9S legumin; in addition, the most abundant monosaccharide is mannose.
The amino acid sequence of α-conglutin is reported in figure1.5.

Legumin-like protein (gi|85361412)

```
  1 illtgnfsrv gfrvdprvre flqyqekegg qgqqqeggnv lsgfddefle ealsvnkeiv
 61 rnikgknddr eggivevkgg lkviipptmr prhgreeeee eeeederrgd rrrrhphhhh
121 heeeeeeeee wshqvrrvrr phrhhhhrkd rngleetlct mklrhniges tspdaynpqa
181 grfktltsid fpilgwlgla aehgsiykna lfvpyynvna nsilyvlngs gmvpscgwqc
241 qcrlqlgtsm rgqvltipte imlql
```

Figure 1.5: Amino acid sequence legumin-like protein of *L. albus*, from NCBI database.

1.2.5 δ-conglutin

Several data in literature indicate that δ-conglutin (a 2S acidic protein) presents different structure and subunit compositions in the 4 lupin species. Regarding *L. albus*, the protein is composed by two chains, called light (MW 14 kDa) and heavy chain (MW 22 kDa). The heavy subunit can be divided in two fractions with similar MW (10.5 kDa), after reduction with β-mercaptoethaol [Restani et al., 1981].
It is characterised by a high level of glutamic acid and glutamine (up to 40%) and, in comparison with the other lupin globulins, it contains more amino acid with acidic residues. The level of glycosylation of the different subunits is very low, and only mannose residues are detected [Duranti et al., 1981]. About *L. angustifolius*, δ-conglutin represents about 20-30% of

the total protein content of lupin seeds, in comparison with the other lupin species in which the δ-conglutin is15-25%. In *L. angustifolius* there are two different isoforms, named δ1 and δ2. The δ1-conglutin is a single subunit of 22 kDa, whereas the δ2-conglutin (the main isoform) is an oligomer composed by different subunits, with a MW ranging from 14 to 22 kDa. Both isoforms, in presence of reducing agents, separate in two disulphide linked chains, called light (MW from 8 to 12 kDa) and heavy chain (MW from 12 to 16 kDa). Glutamic acid and cysteine are the most abundant amino acids, instead tryptophan and methionine are absent [Gayler et al., 1990]. The amino acid sequence of δ-conglutin is reported in figure 1.6

Conglutin delta (gi|80221495)

```
  1 makltilial vaalvlvvht safqsskqsc krqlqqvnlr hcenhiaqri qqqqeeeedh
 61 alklrgikhv ilrhrssqey seeseeldqc ceqlnelnsq rcqcralqqi yesqseqceg
121 sqqeqqleqe leklprtcgf gplrrcdvnp dee
```

Figure 1.6: Amino acid sequence of δ-conglutin of *L. albus*, from NCBI database.

1.3 MASS SPECTROMETRY

Mass spectrometry can be defined as the study of ions in gas phase, and it is an analytical technique measuring the mass-to-charge ratio of ions [Kinter & Sherman, 2000].

This technique has several applications, including:

- identifying unknown compounds by the mass of the compound molecules of their fragments
- determining the isotopic composition of elements in a compound
- determining the structure of a compound in a sample using carefully designed methods (mass spectrometry is not inherently quantitative)
- studying the fundamentals of gas phase ion chemistry (the chemistry of ions and neutrals in vacuum)
- determining other physical, chemical or even biological properties of compounds with a variety of other approaches

Mass spectrometers are composed of three fundamental parts: the ionisation source, the analyser, and the detector (Figure 1.7).

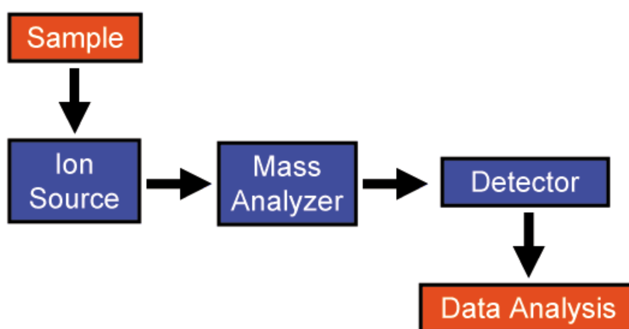


Figure 1.7: A schematic diagram of a mass spectrometer.

The sample has to be introduced into the ionisation source of the instrumental. Once inside the ionisation source, the sample molecules are ionised, because ions are easier to manipulate than neutral molecules. These ions are extracted into the analyser region of the mass spectrometer where they are separated according to their mass (m)- to-change (z) ratios (m/z). The separated ions are detected and this signal sent to a data system where the m/z ratios are stored together with their relative abundance for presentation in the format of an m/z spectrum.
The analyser and detector of the mass spectrometer are maintained under high vacuum to give the ions a reasonable chance of travelling from one end of the instrument to the other, without any hindrance from air molecules. The entire operation of the mass spectrometer, and often the sample introduction process also, is under complete data system control on modern mass spectrometers.
The inlet modes are essentially the high performance liquid chromatography (HPLC) and the direct introduction through infusion pump or a probe.
Many ionisation methods are available and each has its own advantages and disadvantages. The ionisation method to be used basically depends on the type of sample under investigation and includes the following:

- Atmospheric Pressure Chemical Ionisation (APCI)
- Chemical Ionisation (CI)
- Electron Impact (EI)
- Electrospray Ionisation (ESI)
- Fast Atom Bombardment (FAB)
- Field Desorption/ Field Ionisation (FD/FI)
- Matrix Assisted Laser Desorption Ionisation (MALDI)
- Thermospray Ionisation (TSP)

The ionisation methods used for the majority of biochemical analyses are Electrospray Ionisation (ESI) and Matrix Assisted Laser Desorption Ionisation (MALDI).
With most ionisation methods there is the possibility of creating both positively and negatively charged sample ions, depending on the proton affinity of the sample.
Mass analysers separate the ions according to their mass-to-charge ratio. All mass spectrometers are based on dynamics of charged particles in electric and magnetic fields in vacuum.
There are many types of mass analyzers, using either static or dynamic fields, and magnetic or electric fields, but all operate according on the base of mass-to-charge ratio. Each analyzer type has its strengths and weaknesses. Many mass spectrometers use two or more mass analyzers listed below, there are other less common ones designed for special situations. Some of these are sectors, time of flight (TOF), quadrupole, ion trap, Fourier transform ion cyclotron resonance, etc.
Since in this doctoral thesis work it was used an HPLC-Chip-ESI-ion trap instrument, the HPLC-Chip, the electrospray ionisation procedure and the ion trap mass analyser will be discuss in detail.

1.4 HPLC CHIP

The Agilent HPLC-Chip is a laminated polyimide device which is simultaneously at once a nanoelectrospray interface to a mass spectrometer, an analytical LC column (ZORBAX 300SB-C18, 5µm, 75µm, 43 mm length) of a size appropriate to the nanoelectrospray flow rate (0.3 ul/min), and an enrichment column (40 nL ZORBAX 300SB C18) for online sample concentration prior to the analytical column.

In a HPLC-Chip system there are no fittings, adapters, connectors, or any other dispersive elements which are prone to leak and can plague chromatographic performance in capillary nano-LC systems. Postcolumn dispersive effects are exponentially more critical in nanoflow LC than in conventional chromatography which has column volumes on the order of a few milliliters. In fact, the volume of a 75 um id nano-LC column is only a fraction of a microliter. It requires a subnanoliter transfer line to the MS detector in order to preserve the chromatographic separation. The microfabricated HPLC-Chip is a convenient and reliable way to meet such a requirement. Transfer volume between the enrichment column and the analytical column, or other on-chip functions, such as an ion exchange column, is minimized by installing the HPLC-Chip within an LC rotary valve. In one common configuration the HPLC-Chip is interposed between the rotor and stator with precise registration so that in one position flow from the liquid autosampler is directed through an enrichment area, a short packed column of RP material intended to adsorb sample components, and from there to waste (Figure 1.8A). When the rotor travels 60°, the flow from the nanopump enters the enrichment column, sweeping the sample(s) into the analytical column (Figure 1.8B). At the end of the column the flow passes electrical contacts which allow the biasing of the effluent for electrospray. The highly stable spray exits the 2 mm long 40 um od tip and enters the MS analyzer.

The use of this nanotechnology allowed a dramatic increase in sensitivity and reproducibility of the protein detection respect to the use of conventional columns. In proteomics research, identification of proteins depends on identification of peptides which result from enzymatic digests. Vollmer et al. compared HPLC-Chip system and conventional nano-LC system for proteome analysis of yeast [Vollmer et al., 2005]. They found the chip format delivers reduced sample loss and shorter gradient delay time. In the end, more peptides and proteins were identified with the chip system than with traditional capillary nano-LC. In another comparison study, a five-fold increase in sensitivity using the HPLC Chip/ MS system compared to conventional nano-LC system on the same ion trapMS was reported [Hardouin et al., 2006]. Fortier and co-worker. [Fortier et al., 2005] sought to determine the suitability of an HPLC-Chip system for biomarker discovery. The system's ability to accept small dilute samples, its robustness (hundreds of injections), reproducibility in retention time and MS intensity, and superior chromatographic performance were cited as critical for these demanding applications.

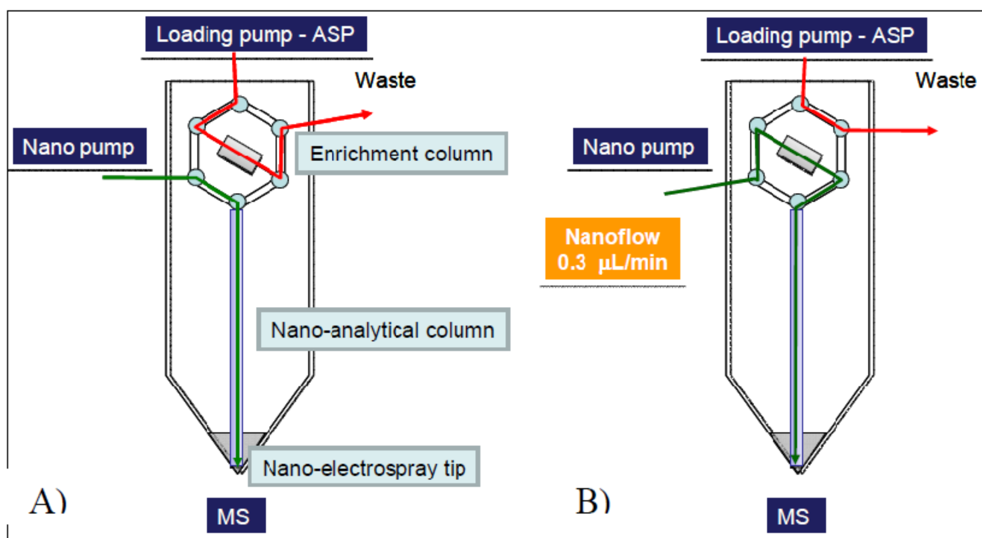


Figure 1.8: Schematic of chip in the loading and running positions.A) The sample loading configuration of the rotorchannels: loading pump loads sample on the enhrichment column at 4ul/min. B) The LC running configuration of the rotor channels. The nano flow (0.3 ul/min) enters the enhrichment column enabling the sample separation on the analytical column. Sampl loading is time dependent and rotor switching is automatic.

1.5 Electrospray ionisation

The phenomena of electrospray has been known for about tens of years, but it was not until the early parts of the 20th century that its significance to science was fully understood [Chapman, 1937]. Some 30 years later, the pioneering experiments by Malcom Dole [Dole et al., 1970] demonstrated the use of electrospray to ionise intact chemical species and the technique of electrospray ionisation (ESI) was invented [Dole et al., 1970]. A further 20 years elapsed until work in the laboratory of John Fenn demonstrated for the first time the use of ESI for the ionisation of high mass biologically important compounds and their subsequent analysis by mass spectrometry [Fenn et al., 1984]. This work was to win John Fenn a share of the 2002 Nobel Prize for chemistry. In the original papers from the late 1980's Fenn and his co-workers successfully demonstrated the basic experimental principles and methodologies of the ESI

technique, including soft ionisation of volatile and thermally labile compounds, multiple charging of proteins and intact ionisation of complexes. ESI-MS is now a basic tool used in probably every biochemistry laboratory in the world.
The analyte is introduced to the source in solution either form a syringe pump or as the eluent flow from liquid chromatography. The analyte solution flow passes through the electrospray needle that has a high potential difference (with respect to the counter electrode) applied to it (typically in the range from 2.5 to 4 kV). This forces the spraying of charged droplets from the needle with a surface charge of the same polarity to the charge on the needle. The droplets are repelled from the needle towards the source sampling come on the counter electrode (shown in blue). As the droplets traverse the space between the needle tip and the cone and solvent evaporation occurs.

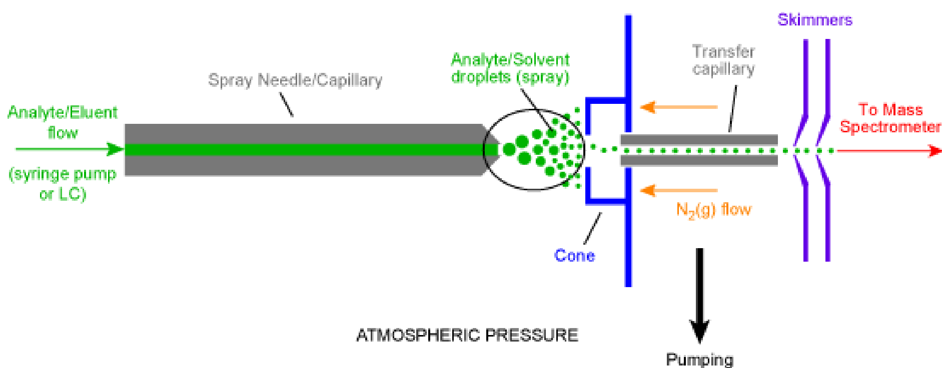


Figure1.9: Schematic representation of electrospray source.

As the solvent evaporation occurs, the droplet shrinks until it reaches the point that the surface tension can no longer sustain the charge (the Rayleigh limit) at which point a "Coulombic explosion" occurs and the droplet is ripped apart. This produces smaller droplets that can repeat the process as well as naked charged analyte molecules.

These charged analyte molecules can be singly or multiply charged. This is a very soft method of ionisation as very little residual energy is retained by the analyte upon ionisation. This is why ESI-MS is such an important technique in biological studies where the analyst often requires that non-covalent molecule-protein or protein-protein interactions are representatively transferred into the gas-phase. The major disadvantage of the technique is that very little fragmentation is produced. For structural elucidation studies, this leads to need for tandem mass spectrometry where the analyte molecules can be fragmented.

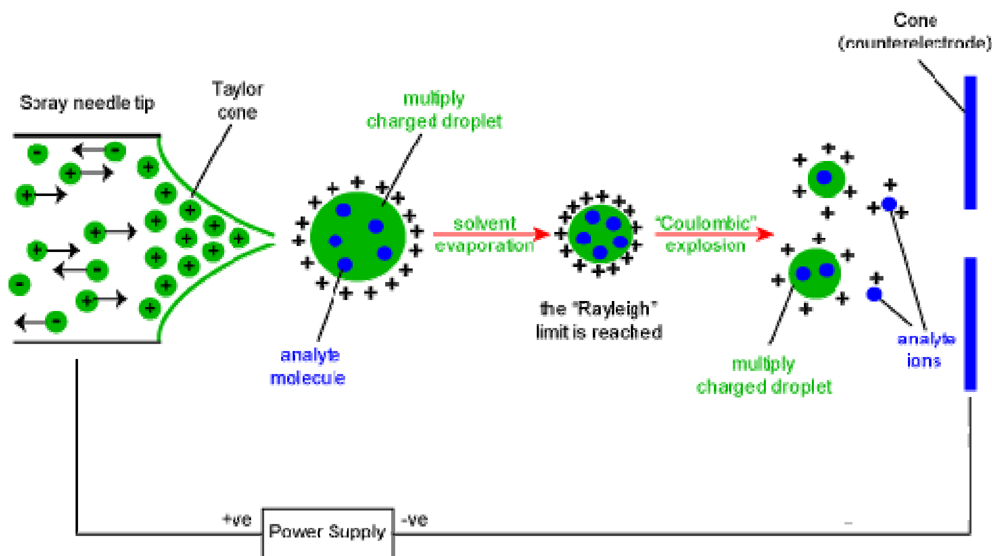


Figure 1.10: Schematic representation of electrospray ion formation.

1.6 Electrospray (ESI) and Nano-Electrospray (nano-ESI)

Nanospray-ESI is a development of ESI for spraying very low amounts of very low concentration samples (nmol/mL). This increased performance is the result of lowering the inner diameter of the spray needle and reducing potentials normally used in ESI. When the flow rate is reduced to nanoliters per minute (nL/min), droplet formation occurs more readily, requiring only the applied voltage to generate spray. No sheath gas or additional heat is required. Consequently, the stability of spray, and therefore signal, at the lower flow rates is typically improved for aqueous or "salty" mobile phases. Nanospray has become a popular method employed in protein analysis. Low flow ESI is especially tolerant to a wide range of liquid compositions, and can even spray pure water with a high degree of stability. The efficiency of ionization improves as the flow rate is lowered because less volume of mobile phase passes through the emitter, producing smaller aerosol droplets. The lower flow rates in a nanospray technique also allow for a longer length of analysis time. This provides ample time to perform novel mass spectrometer scan functions to obtain structural information of an analyte. Nanospray also provides for the direct coupling of nanoscale chromatographic methods, thus signal robbing dilution by a sheath or make-up liquid is eliminated [Covey, 2002].

1.7 Mass analyzer

The mass analyzer is central to mass spectrometric technology, and in the proteomics context, its key parameters are sensitivity, resolution, mass accuracy and ability to produce information-rich fragment mass spectra from peptide ions (tandem mass or MS/MS spectra). There are five basic types of mass analyzers currently used in proteomics: ion trap (IT), Time of-flight (TOF), quadrupole (Q), Fourier transform ion cyclotron resonance (FT-ICR), and the newly developed Orbitrap system [Hu et al., 2005; Yates et al., 2009]. They are different in conception and performance, each with its own strengths and weaknesses. Often, they work as stand-alone mass analyzer, but the current trend points towards hybrid systems in order to combine the advantages of different analyzers in one mass spectrometer: triple-Q, Q-IT, Q-TOF, IT-TOF, TOF-TOF, ITFTICR or IT-Orbitrap tandem mass spectrometers are all capable of protein or peptide sequencing. IT-FT-ICRs and IT-Orbitrap are especially efficient when combined with new fragmentation techniques such as electron capture dissociation (ECD) [Bakhtiar, 2006] or electron-transfer dissociation (ETD) [Wiesner et al., 2008].

1.7.1 Quadrupole ion trap

Ion traps are mass analyzers able to traps ions in a confined space. Because of MS and MS/MS analyses are both performed in the same unit, ion traps are in-time type mass analyzer. This means that steps necessary to obtain the MS/MS spectrum of a precursor ion, i.e. accumulation, isolation and fragmentation, occur sequentially in time but in the same space. Since fragment ions generated from the precursor ion remain trapped as well, it is possible to iterate the fragmentation process forming MS^n spectra. There are three types of ion traps (3D ion trap, 2D ion trap, Orbitrap). 3D ion traps, also called quadrupole ion trap, have been an important mass spectrometer for proteomic experiments.
The trap itself generally consists of two hyperbolic metal electrodes with their foci facing each other and a hyperbolic ring electrode halfway between the other two electrodes. The ions are trapped in the space between these three electrodes by AC and DC (non-oscilling, static) electric fields. The AC radio frequency voltage oscillates between the two hyperbolic metal electrodes at the 'top' and 'bottom' of the trap ('top' and 'bottom' are in phase) and the hyperbolic ring electrode that forms the 'side' of the trap. The ions are first pulled up and down axially while being pushed in radially. The ions are then pulled out radially and pushed in axially (from the top and bottom). In this way the ions move in a complex motion that generally involves the cloud of ions being long and narrow and then short and wide, back and forth, oscillating between the two states.

The quadrupole ion trap has two configurations: the three dimensional form described above and the linear form made of 4 parallel electrodes. The advantage of this design is in its simplicity, but this leaves a particular consistraint on it modelling. To understand how this originates, it is helpful to visualize the linear form. The trap is designed to create a saddle-shaped field to trap a charged ion, but with a quadrupole, this saddle-shaped electric field cannot be rotated about an ion in the centre. It can only ‘flap’ the field up and down. For this reason, the motions of a single ion in the trap are described by the Mathieu Equations. These equations can only be solved numerically, or equivalently by computer simulations.

There are many mass/charge separation and isolation methods but the most commonly used in the mass instability mode in which the RF potential (an alternative radiofrequency) is ramped so that the orbit of ions with a mass $a > b$ are stable, while ions with mass b become unstable and are ejected on the z-axis onto a detector.

Ions may also be ejected by the resonance excitation method, whereby a supplemental oscillary excitation voltage is applied to the end-cap electrodes, and the trapping voltage amplitude and/or excitation voltage frequency is varied to bring ions into a resonance condition in order of their mass/charge ratio [Kinter & Sherman, 2000; Lierberg, 2002; Peng et al., 2001].

In single MS mode, the wall is open and ions enter from the source. Ion traps of recent vintage are equipped with a mechanism to sense the charge density in the trap. When the charge density in the trap reaches a pre-determinated level or a present accumulation time has elapsed, the source is gated to prevent additional ions from entering the ion trap.Unlike other types of mass analysers, the ion trap is kept at a relatively high pressure with helium, which is referred to as the cooling gas. Ions propelled by them source enter the trap with a fair amount of energy. In the absence of the cooling gas the excess of kinetic energy would enable the ions to escape the trap. However, collision with He molecules lowers the ion kinetic energy and brings them under the influence of the trap quadrupole field. The ions contained therein are selectively ejected from the trap based on their mass-to-charge ratio and propelled toward the detector.

In a fragmentation experiment (MS/MS or tandem mass experiment), the trap is configured to accept specified parent ion, in a step that involves the ejection of all the other from the trap. Once the specified charge density is reached, the trap is closed and further ion collection is halted. At this point, a specified fragmentation energy is applied on the end-cap (RF voltage), based on the selected m/z, and the product ions are selectively ejected from the trap and propelled toward the detector. Unlike CID, which delivers the total fragmentation product, sodium adducts and other difficult-to-fragment ions present no problem with the ion trap. Because one can lengthen the accumulation time of the trap, it is possible to select and fragment ions easily even if they are of low abundance.

An ion trap is able to perform three kinds of MS/MS: full scan data-dependent fragmentation, single ion monitoring (SIM) and multiple reaction monitoring (MRM). Working in the data dependent full scan mode, the ion trap performed the MS/MS experiments on the largest possible number of precursor ions: for each scan the ion trap fragments a certain number of precursor ions among the most abundant. So, a subset of high signal peptides seen in the first MS stage (MS1) is subjected to the second MS/MS stage (MS2). In single ion monitoring (SIM) the ion trap isolates and fragments only a precursor ion with a specific m/z ratio. In multiple reactions monitoring (MRM), more than one specific precursor ions (up to ten per time

segment) are selectively monitored along all the analysis. These ions are isolated and fragmented into the ion trap during alternated cycles of MS1 and MS2, which are sequentially repeated for each parent ion. Consequently, the MRM approach provides a high structural specificity for the target precursor ions.
One of the main advantage of the MS/MS experiment is the enhanced specificity. Ion trap analysers has the MS^n capability (up to 11 MS/MS experiments), which is extremely useful in structural elucidation studies, i.e. metabolite characterisation. A subtype of MS/MS experiment is the multiple reaction monitoring (MRM), suitable in quantification analyses of even pmol quantities of analytes.

1.7.2 Linear two dimensional ion trap (LTQ)

Several issues limit the performance of three-dimensional ion traps. First, there is a limit to the number of ions that can be trapped in the device. Second, when ions are scanned from the trap, half exit in the direction of the detector and the other half exit in the opposite direction. Third, there is a limitation in mass accuracy and resolution, although a narrow mass range scan can be employed to obtain high-resolution data with improved mass accuracy. To circumvent some of these limitations, new mass spectrometers based on a two-dimensional quadrupole ion trap mass spectrometer have been developed. Two-dimensional quadrupole ion traps or linear ion traps can hold almost 10 times more ions than three-dimensional traps [Yates et al., 2009]. This increased volume significantly improves ion statistics during mass analysis. Ions are injected into the linear trap through an end cap and then ejected from it. Doing so allows the use of two detectors, since ions exit equally through the sides of the trap. Collecting ions with two detectors doubles the ion current collected during a scan of the m/z range. A second feature of the linear ion trap is the ability to scan at much faster speeds (15,000 AMU/s versus 5500 AMU/s), which increases the number of scans that can be acquired in over the course of an LC analysis. Linear ion traps have limits to the mass resolution or accuracy that can be obtained. At normal scan speeds unit resolution is obtained, but slowing the scan speed can yield much higher resolutions (15,000 resolution over a 10 AMU window). As scan speeds are decreased, the mass range has to be reduced to minimize space charging—a phenomenon resulting when ions of like charge are forced closely together, resulting in a perturbation of ideal ion motion in the electric fields. Tandem mass spectrometry (MS/MS) experiments are performed in this device by separating m/z measurement in time rather than space from ion isolation and collision-induced dissociation. MS/MS experiments benefit from better ion statistics of the linear ion trap as well as increased scan speed. Consequently, more data can be acquired at better quality over a three-dimensional ion trap, but the resolution and mass accuracy measurements of the ion trap are still limited.

1.7.3 LTQ–Orbitrap

A new type of mass analyzer, the Orbitrap, was invented by Makarov in 1999 and was reported as a tool for proteomics research in 2005 by Hu and co-worker. [Hu et al., 2005]. In the Orbitrap, ions are trapped and they orbit around a central spindle-like electrode and oscillate harmonically along its axis with a frequency characteristic of their m/z values, inducing an image current in the outer electrodes that is Fourier transformed into the time domain producing mass spectra. Orbitrap consists of a LTQ coupled to a C-trap and the Orbitrap. It combines the robustness, sensitivity, and MS/MS capability of the LTQ with very high mass accuracy and high resolution capabilities of the Orbitrap, and has become a powerful tool in proteomics. The instrument is capable of mass resolution in excess of 40 000 and mass measurement accuracies of less than 2 ppm for the analysis of complex peptide mixture..

The Orbitrap mass analyzer features high resolution (up to 150,000), high mass accuracy (2– 5 ppm), a mass-to-charge range of 6000, and a dynamic range greater than 103 [Hu et al., 2005; Yates et al., 2009; Bakhtiar et al., 2006; Wiesner et al., 2008; Makarov et al., 2006].

When coupled to an LTQ ion trap, the hybrid instrument has the advantages of both high resolution and mass accuracy of the Orbitrap and the speed and the sensitivity of the LTQ. Furthermore, one can operate LTQ-Orbitrap in a parallel fashion: the Orbitrap acquires MS full scans while the LTQ carries out fragmentation reactions. There are several papers that review and benchmark the performance of the Orbitrap for bottom-up [Perry et al., 2008; Olsen et al., 2005; Yates et al., 2006] and top down [Macek et al., 2006; Frank et al., 2008] proteomic applications.

1.8 Proteomics

Proteomics is one of the most interesting applications of mass spectrometry. The term proteomics means the study of the proteome that is the whole proteins encoded by the genome [Magni et al., 2002]. The genome is the global pool of genes of an organism; some steps are necessary in order to allow that gene information could be expressed as proteome: DNA transcription into an m-RNA sequence, m-RNA translation into an aminoacid sequence (protein precursor), and then post-translational modifications on the precursor in order to give the mature proteins [Kazmi et al., 2001].
The proteome is not stationary, but it can charge during the development of the organism and the environmental conditions. The result of the m-RNA primary transcription could be undergoing to alternative splicing giving several mature m-RNAs codifying for different proteins. In addition, being proteins subjected to post-translational modifications, the number of proteins in proteome will be bigger than that of the gene in genome [Kazmi et al., 2001; Peng et al., 2001].
Among the series of techniques with which proteins can be investigated on a large scale, mass spectrometry (MS) has gained popularity because of its ability to handle the complexities associated with the proteome. Other techniques such as two-dimensional gel electrophoresis (2DE) and protein microarrays fail to achieve the depth of informative proteome analysis seen with MS. The three primary applications of MS to proteomics are cataloguing protein expression, defining protein interactions, and identifying sites of protein modification. The use of MS for proteomics is not the application of a single technique for all purposes but rather a collection of methodologies, each with strengths suited to particular inquiries. For any MS experiment, consideration should be given to the type of instrumentation, fragmentation method, and analysis strategy best suited to an individual sample and to a goal predetermined.
Different strategies for MS-based protein identification and characterization were described.
Proteins extracted from biological samples can be analyzed by top down or bottom-up methods. In the top-down approach, a whole-protein analysis is performed. Proteins in complex mixtures are fractionated and separated into pure single proteins or less complex protein mixtures, followed by infusion of sample into the mass spectrometer for intact protein mass measurement and/or intact protein fragmentation [McLafferty et al., 2007]. In the bottom-up approach, proteins in complex mixtures can be separated before enzymatic digestion followed by direct peptide mass fingerprinting-based acquisition (MALDI-TOF) or further peptide separation on-line coupled to tandem mass spectrometry (liquid-chromatography coupled with mass spectrometry, LC-MS/MS). Alternatively, the protein mixture can be directly digested into a collection of peptides (‘shotgun’ proteomics), which are then separated by monodimensional or multidimensional chromatography on-line coupled to tandem mass spectrometric analysis (LC-MS/MS or 2DLC-MS/MS) [Wu et al., 2002].

1.9 Shotgun proteomics

Multiple strategies have been developed to systematically and comprehensively profile biological systems. 'Shotgun proteomics' refers to the direct and rapid analysis of the entire protein complement within a complex protein mixture. Implicit in this methodology is the ability to monitor the system both qualitatively and quantitatively. In fact, a comprehensive proteomic analysis should ideally include the following functions: a) identify the entire protein complement; b) detect post-translational modifications (PTM); and c) allowing for quantitative comparisons between samples. Bottom-up strategies, in which peptide detection is used to infer protein presence, are the standard for large-scale or high-throughput analysis of highly complex samples such as direct tissue proteomics [Rezaul et al., 2008; Hwang et al., 2007]. Shotgun proteomics relies on the digestion of protein mixtures followed by separation of the peptides and subsequent introduction into a tandem mass spectrometer. Tandem mass spectra are collected for as many peptides as possible, and the results are then searched by an algorithmic comparison, via Sequest or Mascot for example, against a database of proteins derived from genomic sequencing to identify the peptides (Figure 1.11). These approaches are well suited to the analysis of protein complexes, which comprise a discrete set of proteins with a functional relationship, and shotgun proteomics provides a sensitive technique to identify the components of complexes. These are usually conducted in two workflows. ‘Sort-then-break’ approaches are performed using off-line protein fractionation and separation before protein digestion, followed by direct peptide analysis by ‘peptide mass fingerprinting’ (PMF) or further peptide separation by LC interfaced to a tandem mass spectrometer. In ‘break-thensort’ approaches, protein digestion is performed without any prefractionation/separation and peptides are separated by multidimensional chromatography followed by tandem mass spectrometric analysis, typically using rapidly scanning analyzers such as ion trap (IT) mass spectrometers.

The resolution and peak capacity of the separation techniques coupled with tandem mass spectrometry, in particular of the liquid chromatographic system in LC-MS/MS, are crucial to the success of the analysis. Although the shotgun approach is conceptually simple, it requires highly sensitive and efficient separation. Information is also lost upon the conversion of intact proteins into a mixture of peptides, which can lead to incorrect identifications. Not all peptides resulting from the digestion of a protein can be observed or correctly identified with MS analysis, especially those with unexpected modifications. Furthermore, the limited dynamic range of mass spectrometric analysis only allows for the peptides present at high relative abundance to be preferentially sampled, while information regarding the proteins represented as low abundance peptides in the complex mixture may be not obtained.

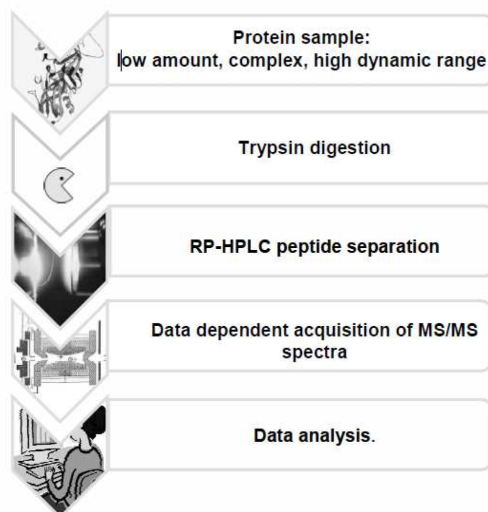


Figure 1.11: Workflow of the bottom-up approach "shotgun proteomics". Shotgun proteomics enables the identification of a large number of proteins in high complex samples, characterized by a high dynamic range, without previous fractionation. The complex mixture of proteins is digested with trypsin to obtain a more complex mixture of peptides. Peptides were separated on the basis of their hidrophobicity using a reverse phase (RP) liquid chromatographic separation techniques (high pressure liquid chromatography, HPLC) coupled with tandem mass spectrometry. The acquisition of the fragmentation spectra (MS/MS spectra) of peptides occurs in a data dependent acquisition mode. Thus means that as many peptides as possible are fragmented into the mass analyzer (ion trap) during the LC-MSMS analysis with the aim to identify as many protein as possible in the complex sample. The identification of the protein is done using opportune software which are able to compare the experimental MS/MS spectra with theoretical fragmentation spectra deduced by the in-silico digestion of sequences stored in a database.

1.10 Mass spectrometric instrumentation for shotgun proteomics

MS measures the mass-to-charge ratio (m/z) of gas-phase ions. Mass spectrometers consist of an ion source that converts molecules into gas-phase ions, a mass analyzer that separates ionized analytes on the basis of m/z ratio, and a detector that records the number of ions at each m/z value. The development of electrospray ionization (ESI) [Liebler et al., 2002; Kinter et al., 2000] and matrix-assisted-laser desorption/ionization (MALDI) [Liebler et al., 2002; Kinter et al., 2000], the two soft ionization techniques capable of ionizing peptides or proteins, revolutionized protein analysis using MS. Both MALDI and ESI are soft ionisation techniques in which ions are created with low internal energies and thus undergo little fragmentation in source. In MALDI, samples are co-crystallised with an organic matrix on a metal target. A pulsed laser is used to excite the matrix, which causes rapid thermal heating of the molecules and eventually desorption of ions into the gas phase. Because of the usage of a pulsed laser, MALDI produces packets of ions rather than a continuous beam; it is therefore most often coupled to a mass analyzer that can measure either a complete mass spectrum without scanning a mass range, or trap all the ions for subsequent mass analysis like time-of-flight (TOF) mass analyzer. This ionisation technique tolerates a reasonable amount of impurities in the sample to be analysed.

1.11 Fragmentation techniques

Tandem mass spectrometry (MS/MS) is a key technique for protein or peptide sequencing and post translational modification (PTM) analysis. Collision-induced dissociation (CID) has been the most widely used MS/MS technique in proteomics research. In this method, gas-phase peptide/protein cations are internally heated by multiple collisions with rare gas atoms. This leads to peptide backbone fragmentation of the C–N bond resulting in a series of b-fragment and y-fragment ions. Because of the slow-heating, energetic feature associated with this method, the internal fragmentation and neutral-losses of H_2O, NH_3, and labile PTMs are common. This also results in limited sequence information for large peptides (>15 amino acids) and intact proteins.
All information necessary to rebuild the primary sequence of a peptide is comprised in the MS/MS spectrum of the precursor peptide ion. There are three different types of bonds that can fragment along the amino acid backbone: the NH-CH, CH-CO, and CO-NH bonds. Each bond cleavage gives rise to two species, one neutral and the other one charged, and only the charged species is monitored by the mass spectrometer.

The charge can stay on either of the two fragments depending on the chemistry and relative proton affinity of the two species. Hence there are six possible fragment ions for each amino acid residue and these are labelled with the a, b, and c ions having the charge retained on the N-terminal fragment, and the x, y, and z ions having the charge retained on the C- terminal fragment. Trypsin digestion is very suitable for mass spectrometric studies because each proteolytic fragment contains a basic arginine (R) or lysine (K) amino acid residue, and thus is eminently useful for positive ionisation mass spectrometric analysis.

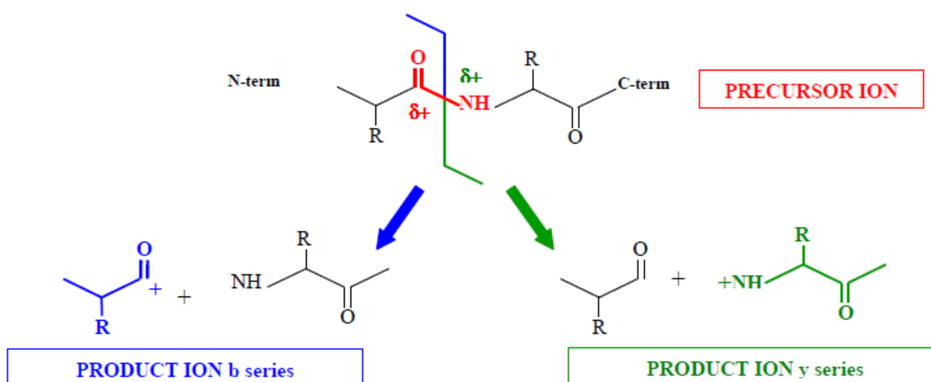


Figure 1.12: Main product ion types (y serie and b serie) obtained in CID fragmentation.

Using low energy fragmentations (CID), the most common cleavage sites are at the CO-NH bonds which give rise to the b and/or the y ions (Figure 1.12).

The mass difference between two adjacent b ions, or y; ions, is indicative of a particular amino acid residue. A consequence of the low energy involved during fragmentation is the detection of fewer types of side-chain fragments.

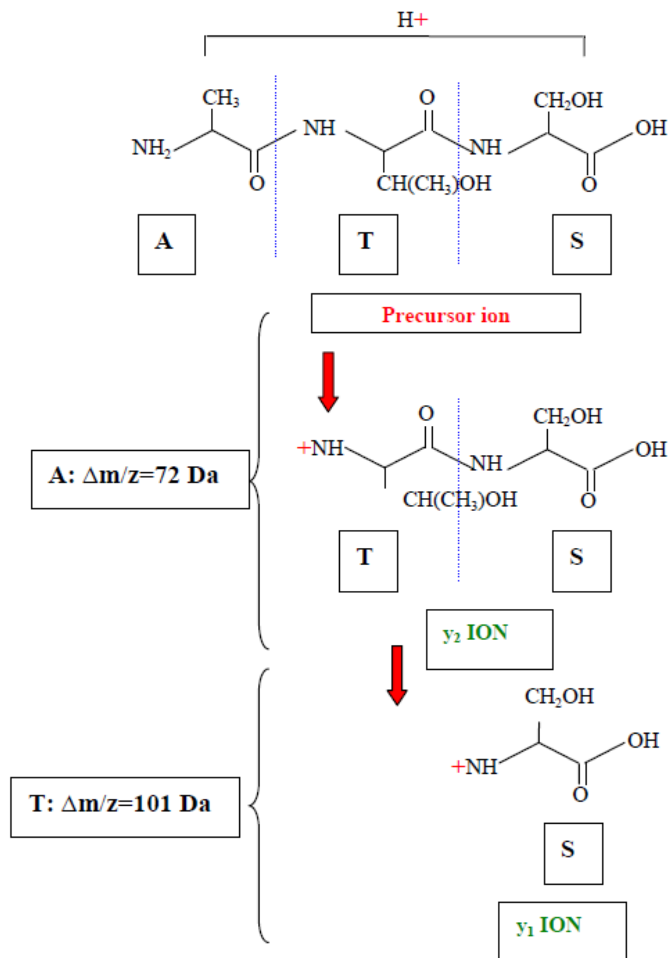


Figure 1.13: Schematic representation of the sequencing theory from tandem mass spectrometry. The variation between the m/z ratio between two adjacent ion products belonging to the same series corresponds to a specific aminoacid residue.

Immonium ions (labelled “i”) are formed by losing a molecule of CO, and they appear in the very low m/z range of the MS/MS spectrum. Each amino acid residue leads to a diagnostic

immonium ion, with the exception of the two pairs leucine (L) and iso-leucine (I), and lysine (K) and glutamine (Q), which produces immonium ions with the same m/z ratio, i.e m/z 86 for I and L, m/z 101 for K and Q. The immonium ions are useful for detecting and confirming many of the amino acid residues in a peptide, although no information regarding the position of these amino acid residues in the peptide sequence can be ascertained from the immonium ions.

Varying amounts of sequence information can be gleaned from each fragmentation spectrum, and the spectra need to be interpreted carefully. Manual interpretation of an MS/MS spectrum is a very time and energy consuming process; because of this, much powerful software have been introduced with the aim of data processing and protein identification against databases.

These softwares (Sequest, Mascot, Spectrum Mill), starting from the unprocessed analyses rebuild the amino acid sequence, and search against the whole databases identifying statistically the target protein.

A new fragmentation technique, electron-capture dissociation (ECD) was introduced by the McLafferty laboratory in 1998 by which the capture of a thermal electron by a multiply protonated peptide/ protein cation induces backbone fragmentation at the N–Cα bond to produce c-type and z-type fragment ions. ECD provides more extensive fragmentation resulting in richer MS/MS spectra and better sequence coverage, and the nonergodic feature of ECD preserves labile PTMs. Therefore, it has become a powerful tool for top-down analysis of intact proteins [Ge et al., 2002]. However, ECD is most often constrained to the expensive, highly sophisticated FTICR instruments.

An analogous technique, electron-transfer dissociation (ETD) was developed by the Hunt laboratory in 2004 and extends electron-capture-like fragmentation to more common bench top mass spectrometers [Udeshi et al., 2008]. In this process, the electrons transfer from radical anions with low electron affinity to multiply protonated peptide cations initiating backbone fragmentation to produce c ion and z-ion series. Because the ion/ion reaction is highly efficient and fast, ETD can easily be performed with femtomole quantities of peptides on a chromatographic timescale. ETD MS/MS provides superior sequence coverage for small-sized to medium-sized peptides and is highly complementary to conventional CID for proteome identification applications [Mikesh et al., 2006]. ETD can be utilized to analyze very large peptides as well as intact proteins with a sequential ion/ion reaction, proton transfer/ charge reduction (PTR) by which the ETD produced multiply charged fragments are deprotonated with even electron anions resulting in singly and doubly charged ions that are readily measured by the bench-top low resolution ion trap instrument. This allows for the sequence analysis of 15–40 amino acids at both N-terminus and C-terminus of the protein. ETD has also shown great promise in labile PTM analyses such as phosphorylation [Chi et al., 2007].

1.12 Quantitative proteomics

1.12.1 Relative quantitative proteomics

The mere identification of a protein expressed in a biological system is not sufficient to answer most biological questions because quantitative answers are more and more required providing a snapshot of the protein expression state of a cell in response to biological perturbations such as cell morphogenesis, disease progression, or drug treatment.
This requires sensitive and accurate assays for identifying proteins in complex mixtures and quantifying their abundances. While still producing excellent results, "gel-based" quantitative proteomics has been largely superseded by "gel-free" MS-based quantitative proteomics approaches where quantification is performed using the mass spectrometric data. Gel-based quantitative proteomics is limited in sensitivity and can be inefficient when analyzing insoluble proteins or those with very high or low mass and pI value.
Similar to the gel approach, where per se the protein staining intensity within a gel is not proportional to the amount present within sample, in both MALDI and ESI-MS the relationship between the amount of protein present and the measured signal intensity is complex. Also the reproducibility of a peptide/protein signal between different runs is complex. Therefore, a great knowledge of LC-MS features is an essential need for the development of reliable quantitative methods and, moreover, opportune techniques to alleviate mass-spectrometry related problems in quantitative approaches were developed.

1.12.1.1 Stable Isotope Labeling techniques

Most of the quantitative proteomics approaches by MS utilize isotopic labels as a reference for relative quantification (Stable-Isotope Labeling techniques). This method makes use of the fact that pairs of chemically identical molecules (in this case peptide pairs), but with different stable-isotope composition (^{13}C instead of ^{12}C, ^{2}H instead of ^{1}H, ^{18}O instead of ^{16}O or ^{15}N instead of ^{14}N) can be differentiated in a mass spectrometer owing to their mass difference only. Thus the ratio of signal intensities for such peptide pairs should be a direct and accurate measure of the abundance ratio between the two peptides/proteins derived from two different biological conditions. Three main approaches exist today, which are: a) metabolic stable isotope

labeling, b) isotope tagging by chemical reaction and, c) stable-isotope incorporation via enzyme reaction.

The main metabolic method is stable-isotope labeling by amino acids in cell culture (SILAC) [Ong et al., 2002], in which amino acids containing stable isotopes, like arginine with six 13C atoms, are supplied in growth media. Several amino acids have been used like leucine (deuterated form), which labels 70% of tryptic peptides [Foster et al., 2003], or simultaneously lysine and arginine, with subsequent tryptic digestion resulting in labeling of all peptides but the C-terminal peptide [Ibarrola et al., 2003]. A principal advantage of metabolic labeling over chemical labeling is the earliest possible introduction of the label into the live cells, immediate pooling of case and control and the concomitant reduction of parallel sample preparation bias. The absence of "harsh" chemistry and side reactions is also an advantage. While these methods can only be applied to cultured cells like bacteria or yeast, recently these organisms have in turn been fed to small multicellular organisms such as Caeonorhabditis elegans, Drosophila melanogaster [Krijgsveld et al., 2003], plants [Ippel et al., 2004] or even a rat by using 15N labeled algae [Wu et al., 2004]. Even more promising is the pairwise comparison between cultured cell lines and dissected tissues [Ishihama et al., 2005]. In this case, a cell line derived from the tissue in question is labeled with SILAC and then spiked into both tissue states (e.g. healthy vs. diseased tissue) to serve as an internal standard and independent reference for both conditions. Thus, if the two ratios (healthy tissue vs. internal standard and diseased tissue vs. internal standard) obtained with the internal standard are different, it directly reflects a change in protein expression between compared tissues.

A wide variety of isotopically labeled chemicals has been reported. All chemical reagents are targeted toward reactive sites on a protein or peptide and the two proteomes to be compared are labeled with the light and heavy reagent, respectively. Isotope-coded affinity tagging (ICAT) [Gygi et al., 1999] was the first approach described in 1999 by Gygi and co-workers. This agent consists of a reactive group that is cystein-directed, apolyether linker region with eight deuteria and a biotin group for avidin purification of labeled peptides. Due to compromised co-elution of deuterium-tagged and natural hydrogen peptides, and MS fragmentation problems (large tag) with this first ICAT version, a new version was developed with an acid cleavable site and ^{13}C atoms instead.

Recently, Gygi and colleagues have described a new method called catch-andrelease (CAR) [Gartner et al., 2007] that makes use of a cystein-directed reductively cleavable reagent. The tag features a novel disulfide moiety that links biotin and a thiol-reactive group. The disulfide is resistant to reductive conditions during labeling but readily cleaved with tris-(2-carboxyethyl) phosphine (TCEP), therefore simplifying sample handling procedure sand reducing non-specific interactions during avidin purification. Several strategies have been reported that target amines of which two have been applied to experimental biology. The first, isotope-coded protein labeling (ICPL) [Schmidt et al., 2005], targets all amino groups at the protein level using nicotinoyl oxysuccinimide (Nic-NHS) as the reagent. The second, isobaric tag for relative and absolute quantification (iTRAQ) [Ross et al., 2004], uses the same NHS chemistry as ICPL, but adds an innovative concept, namely a tag that generates a specific reporter ion for quantification in MS/MS spectra (mass 114, 115, 116, 117) but with isobaric mass at MS level. Therefore, mass spectra are relatively simple and differential behavior is only reported after fragmentation.

Moreover, multiplexing (currently eight plex) is an interesting feature as it allows comparing more than two conditions. Carboxylic groups have also been labeled using either methyl [Goodlett et al., 2001] or ethyl [Syka et al., 2004] esterification at the peptide level. However, both methods use deuterium atoms and bear the risk of chromatographic discrimination and the mass offsets of 2Da (methyl) and 4 Da(ethyl) poses problems of isotopic overlap of the peptide pairs. A clear advantage of all these chemical approaches is the multitude of available functional groups in proteins allowing designing almost any kind of quantitative tag. Possible enrichment is also an asset as it allows reducing sample complexity without loosing quantitative information. However, reactions have to be specific, proceed to completion and involve minimal sample handling. Side reactions are problematic, too, as they considerably increase the sample complexity. Despite these constraints, chemical stable-isotope labeling has produced most of the quantitative proteome data mainly due its chemical versatility and certainly because of its applicability to any biological sample as opposed to metabolic labeling.
Stable isotopes can also be introduced into the peptide by different proteases such as trypsin, Lys-Nor Glu-C [Mirgorodskaya et al. 2000; Yao et al., 2001; Rao et al., 2005]. The digestion is performed in $H_2{}^{18}O$ water and enzymatic oxygen exchange occurs at the carboxyl group of the generated peptides. The advantage of this method is its versatility (virtually any protease-generated peptide is labeled), its applicability to low sample amounts and almost unlimited compatibility with sample preparations. On the other hand, the labeling is performed only at peptide level, and samples have to be processed in parallel until these peptides are generated. One or two oxygens can be exchanged leading to variability in peptide spacing and the mass offset of 2 Da is not sufficient to separate the isotopic envelopes. Recent modifications such as post digestion incubation of peptides in small volumes of $H_2{}^{18}O$ or deactivating the protease through reduction/alkylation have addressed these issues [Bantscheff et al., 2004; Staes et al. 2004; Miyagi et al., 2007].

1.12.1.2 Stable isotope label free techniques

Recently, new promising approaches described as “label-free” that do not use labeling and stable isotopes to obtain quantitative information have emerged. Literature proposes different parameters which can be derived from raw MS/MS data to measure protein abundances.

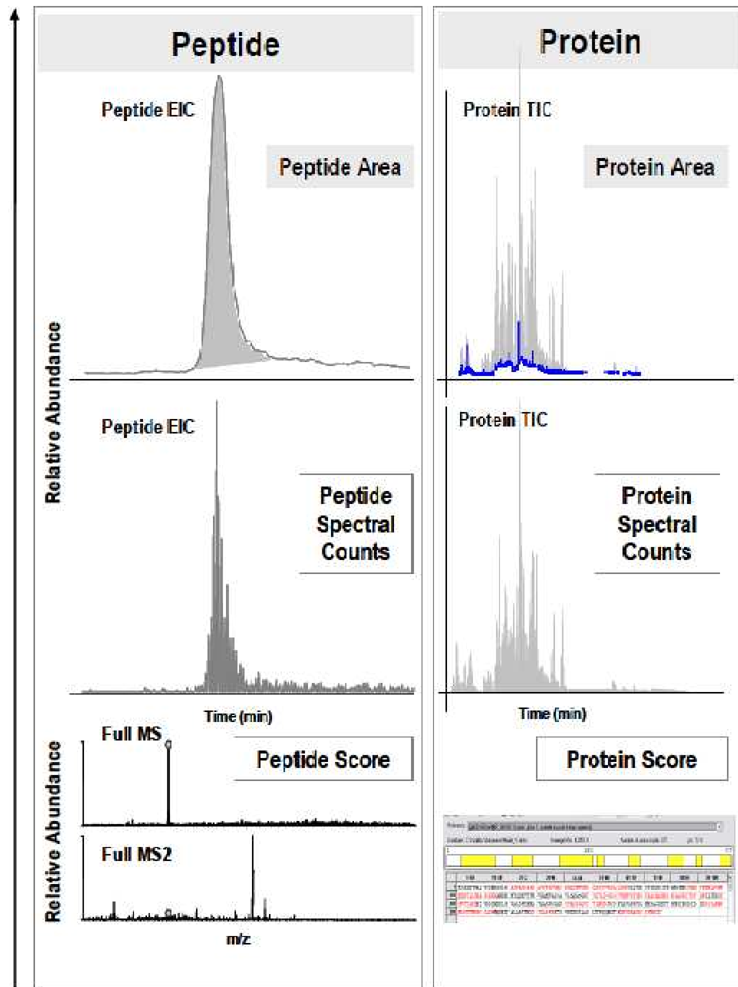


Figure 1.14: Overview and description of the main parameters for the label-free quantitative proteomics. In bottom-up LC-MS/MS a protein is identified thanks to the sequencing of its tryptic peptides starting from their experimental MS/MS spectra which are matched vs. theoretical MS/MS spectra in a database. Different softwares are able to perform the match. All softwares are able to attribute to each MS/MS spectrum a peptide score which is an index of the quality of the match. By summing the scores attributed to all peptides of a certain protein is possible to calculate the "protein score". In a LC-MS/MS run, peptides are associated with more than a single spectrum. The sum of all MS/MS spectra of a peptide is the "peptide spectral counts". The sum of all MS/MS spectra of all peptides attributed to a protein is the "protein spectral counts". In an LC-MS/MS run, it is possible to associate to each peptide an extracted ion chromatogram (EIC) and a correspondent chromatographic peak. The area of the peak is called "peptide area" and is used to calculate the "protein area" which may be defined as the sum of all areas of all peptides attributed to the protein. The complexity of label-free parameters increases moving from protein score to protein area.

A first strategy is based on peptide score summation (PMSS) [Allet et al., 2004] in order to obtain protein score. The method is based on the assumption that a protein score is a sum of identification scores of its peptides and that a high protein score is correlated with a higher abundance, thus yielding semi-quantitative information. The main limitation of protein score depends on its tendency to saturation: each protein sequence showed a limited number of tryptic peptides suitable for mass spectrometry detectability. Therefore, a limit in the increase of protein score exists for each protein.

Another label-free method, termed spectral counting or spectrum sampling (SpS), compares the number of MS/MS spectra assigned to each protein. The spectral counting is the sum of the all MS/MS observations for any peptide in a given protein, including spectra redundant for ion charge states. Spectral counting of standard proteins added to yeast extracts showed linearity over 2 orders of magnitude with high correlation to the relative protein concentration [Liu et al., 2004]. An advantage of spectral counting is that relative abundances of different proteins can in principle be measured. Thus, significant correlations have been shown between spectral counts and independent estimates of protein copy number in yeast [Ghaemmaghami et al., 2003]. Protein abundance indices (PAIs) represent another related method and are believed to be more reliable as they are based on observable parameters. For example, the number of peptides identifying a protein increases with increasing protein amount. As a larger protein will statistically generate more measurable peptides than a smaller one, a simple PAI can be derived by normalizing the number of observed peptides with the number of observable peptides for the protein under consideration [Rappsilber et al., 2002; Sanders et al., 2002]. Ishihama and colleagues have described an exponentially modified PAI (emPAI) by observing a logarithmic relationship between the number of peptides observed and the protein amount within given sample [Ishihama et al., 2005].

Ion intensity based label-free methods are based on the integration of peptide parent ion chromatogram peaks (EICs). They rely on the observation that the peak intensity (or better: peak volume or peptide area) in most cases is proportional to the concentration of the peptide in the sample [Roy et al., 2004; Silva et al., 2004; Schmidt et al., 2003; Shen et al., 2002]. Unfortunately, it is not possible to predict the MS detect or response to a particular peptide because of unknown extraction and peptide ionization properties and, therefore, extracted ion currents (XICs or EICs) from different peptides of the same protein are also very different even if they are present at the same concentration. Although directly comparing intensities between different peptides is not possible for the reasons previously mentioned. These sources of error do not apply when comparing the same peptide in different chromatographic runs using identical experimental conditions. Thus two proteomes can be compared when analyzed one after the other and in exactly the same way [Chelius et al., 2002; Lasonder et al., 2002]. A clear advantage of such method is the absence of any label and the applicability to any type of instrument. Clear disadvantages are the multiple occasions for quantification error to occur during parallel sample processing, analysis and the need for very accurate and reproducible LC and MS runs.

Measurements of mass spectral peak intensities and spectral counts are probably most promising methods for quantifying protein abundance changes in shotgun proteomic analyses.

Peak intensity values useful for protein quantitation ranged from 107 to 1011 counts with no obvious saturation effect, and proteins in replicate samples showed variations of less than 2-fold within the 95% range when >3 peptides/protein were shared between samples [Wang et al., 2003].
Overall spectral counting proved to be a more sensitive method for detecting proteins that undergo changes in abundance, whereas peak area intensity measurements yielded more accurate estimates of protein ratios [Wang et al., 2003].
Three studies using standards have demonstrated that mass spectral peak intensities of peptide ions correlate well with protein abundances in complex samples. Bondarenko and co-workers demonstrated linear responses of peptide ion peak areas between 10 and 1,000 fmol of myoglobin spiked into human plasma with a relative standard deviation <11% [Bondarenko et al., 2002; Chelius et al., 2002]. Wang and co-worker [Wang et al., 2003] published similar results with protein standards spiked into serum, obtaining a median relative standard deviation of 26% for peak intensity ratios from 3,400 ions in 25 replicate measurements.

1.12.2 Absolute quantitative proteomics

The main absolute quantitative methodologies (AQUA, QconCAT) in proteomics rely on the addition of isotope-labeled proteotypic peptides from the target protein(s) to the tryptic digest of the samples. The AQUA method has been successfully applied to the quantification of neuropeptides [Wei et al., 2006] or protein phosphorylation using phosphopeptides as standards [Gerber et al., 2003, Kirkpatrick et al., 2005; Stemmann et al., 2001]. This methodology, however, is very expensive since it requires the individual synthesis, purification, and quantification of all isotope-labeled peptides. For this reason, target proteins are generally quantified using a single AQUA peptide [Barnidge et al., 2004; Cheng et al., 2006], although an accurate quantification can only rarely be based on one single peptide. Recently, artificial concatamers of standard isotope-labeled peptides (QconCAT) [Anderson & Hunter et al., 2006; Beynon et al., 2005; Pratt et al., 2006; Rivers et al., 2007] have been introduced as a smart strategy to achieve multiplex absolute quantification in a single experiment, since several proteotypic peptides, representing either a single protein or different proteins, can be included in the QconCAT.
Although, AQUA and QconCAT have significantly improved the absolute quantitative measurement of proteins in biological samples, the calibration with AQUA peptides and QconCAT constructs have some limitations, such as the poor protein sequence coverage (limiting the statistical reliability of the quantification), the failure to take into account the actual efficiency of the proteolysis step and an incompatibility with sample fractionation, which is often necessary when dealing with biological samples [Shen et al., 2005].
For the absolute quantitative proteomics a two step strategy is required: the first step is the discovery phase in which the proteome, observable by mass spectrometry-based shotgun

proteomics, is extensively analyzed with the aim to identify proteotypic peptides [Anderson & Hunter et al., 2006] related to the protein(s) to be quantified. Proteotypic peptides were characterized by their uniqueness for a single protein and their detectability in mass spectrometry. In the second step, proteotypic peptides are synthesized according to AQUA and used as internal standards for the absolute quantification of target protein.

Anderson and Aebersold [Anderson & Hunter et al., 2006; Mallick et al., 2007] have recently proposed an innovative multiple reaction monitoring (MRM) strategy for protein quantification focused on a defined proteome subset and based on proteotypic peptides (PTPs). This strategy is highly sensible and specific for a target set of proteins and is opposed to the classical "shotgun" way of identification and quantification of as many proteins as possible. The proteome, observable by shotgun proteomics, is extensively analyzed in order to identify and select the best proteotypic peptides belonging to the protein(s) to precisely quantify and/or validate. Then proteotypic peptides are used to obtain sensitive, robust and reproducible measurements based on targeted MRM mass spectrometry analysis. While being less comprehensive than shotgun proteomics, the MRM-PTP strategy appears to be more sensitive and may be used both in relative quantitative proteomics and in absolute quantitative proteomics to validate results of differential analysis which may be obtained using SIL (stable isotope labeling) or SIF (stable isotope label-free) techniques.

1.13 Quantitative proteomics and nutraceutics

The recent application of innovative proteomic tools for the assessment of food quality has revealed its efficiency in pointing out differences in food proteomes relevant for human nutrition [Carbonaro, 2004]. In particular, quantitative proteomics based on MS has been used for detecting and quantifying allergenic proteins or bioactive compounds, including peptides produced during in vivo processing.

In food science and nutrition there is a growing demand of methods able to absolute quantify allergens, toxins, or bioactive proteins. Proteomics has matured from a basic research platform to an analytical tool now widely employed in medicine, pharmacology, biology, and, more recently, also in nutrition and food research. Whereas in the clinics and in biology proteomics is mainly expected to deliver disease biomarkers and drug targets, in food science it is applied to the assessment of the quality and safety of foods or specific food ingredients, in general having as a main objective the quantification of bioactive proteins, in particular allergens, toxins, and nutraceutics [Schetsrieder & Baeuerlein, 2009]. These last are used as bioactive ingredients in the formulation of dietary supplements or functional foods, i.e. foods that beyond adequate nutritional qualities should either improve the state of health and well-being and/or reduce the risk of disease. Plant proteins have an important role in the diet since they have been shown to induce a significant reduction of cholesterolemia both in experimental animal models and hypercholesterolemic humans [Sirtori et al., 2009]. The cholesterol-lowering effect, potentially leading to a reduced cardiovascular risk, was the basis for the U.S. Food and Drug Administration (US FDA) approval of the health claim concerning the role of soybean proteins in reducing the risk of coronary disease. This claim affirms that the consumption of 25 g of soy protein per day is useful in the prevention of cardiovascular diseases (CVDs); because of this, soy proteins have become a successful ingredient in the preparation of functional foods for the prevention of CVDs. All these facts have stimulated researchers on other legume species in order to select additional sources of bio-functional ingredients.

Another functional ingredient that only recently has attracted the interest of research is lupin protein: experimental and clinical investigations have indicated that this protein may be useful for controlling hypercholesterolemia [Bettzieche et al., 2008; Martins et al., 2005; Hall et al., 2005; Sirtori et al., 2004], hyperglycemia [Magni et al., 2004], and hypertension [Pilvi et al., 2005]. The use of the lupin in human nutrition is steadily increasing, since it has favorable nutritional and technological characteristics, which permit to use especially its protein component as ingredient in the formulation of a large range of different food products.

Recent investigations have also pointed out that lupin proteins have potential nutraceutical properties. Studies on established animal models have demonstrated that they are hypocholesterolemic in rats [Bettzieche et al., 2008] and pigs [Martins et al., 2005], anti-atherosclerotic in rabbits [Marchesi et al., 2008], hypoglycemic in rats [Magni et al., 2004], and anti-hypertensive in mice [Pilvi et al., 2005].

Meanwhile, however, some papers have indicated a possible cross-allergenicity with peanut proteins [Magni et al., 2005], which prompted the European Commission to include this seed in the list of food allergens whose declaration in food labels is compulsory. About 90% of the lupin seed proteins are globulins that have the main physiological role of being storage proteins. Out of them, the most relevant in human nutrition are the vicilins (named also β-conglutin) and γ-conglutin. The former have been hypothesized to be the hypocholesterolemic component of lupin proteins for their high homology with the alpha' subunit of soy β-conglycinin [Wait et al., 2005], the major bioactive component of soy proteins, whereas the latter is both hypoglycemic [Magni et al., 2004] and one of the major lupin allergens [Holden et al., 2008]. This last hypothesis is supported by its thermal stability and resistance to proteolysis that are important chemical features frequently shared by food allergens. All these facts indicate that the detection and the quantification of the vicilins and γ-conglutin are very crucial aspects in the analysis of lupin seed and products. In contrast to other leguminous plants (peas, soy, beans), lupins contain extremely low amounts of trypsin inhibitors, lectins, isoflavones, saponins and cyanogens.

1.14 References

Allet N et al., 2004, In vitro and in silico processes to identify differentially expressed proteins. Proteomics 4:2333-2351.

Anderson JW et al., 1995, Meta-analysis of the effects of soy protein intake on serum lipids. New Eng J Med 333:276-282.

Anderson L, Hunter CL, 2006, Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol. Cell. Proteomics 5:573-588.

Bakhit R et al., 1994, Intake of 25 g of soybean protein with or without soybean fiber alters plasma lipids in men with elevated cholesterol concentrations. J Nutr 124:213-222.

Bakhtiar R, Guan Z, 2006, Electron capture dissociation mass spectrometry in characterization of peptides and proteins. Biotechnol Lett 28:1047-1059.

Bantscheff M et al., 2004, Femtomol sensitivity post-digest (18)O labeling for relative quantification of differential protein complex composition. Rapid Commun Mass Spectrom 18:869-876.

Barnidge DR et al., 2004, Absolute Quantification of the Model Biomarker Prostate-Specific Antigen in Serum by LC-MS/MS Using Protein Cleavage and Isotope Dilution Mass Spectrometry. J Proteome Res 3:644-652.

Bettzieche A et al., 2008, Differing effect of protein isolates from different cultivars of blue lupin on plasma lipoproteins of hypercholesterolemic rats. Biosci Biotechnol Biochem 72:3114-3121.

Beynon RJ et al., 2005, Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. Nat Methods 2:587-589.

Blagrove RJ, Gillespie JM, 1975, Isolation, purification and characterization of the globulins of *L. angustifolius*. Australian J Plant Physiol 2:13-27.

Blagrove RJ et al., 1980, Physiochemical studies on conglutin γ a storage globulin from seeds of *Lupinus angustifolius*. Aust J Plant Physiol 7:1-13.

Bondarenko PV et al., 2002, Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. Anal Chem 74:4741-4749.

Borrato K, Dure III L, 1987, The globulin seed storage proteins of flowering plants are derived from two ancestral genes. Plant Mol Biol 8:115-131.

Carbonaro M, 2004 Proteomics: Present and future in food quality evaluation. Trends Food Technol 15:209-216.

Chapman S, 1937, Carrier-mobility spectra of spray-electrified liquids. Physical Review 52:184-190.

Chelius D, Bondarenko PV, 2002, Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. J Proteome Res 1:317-323.

Cheng D et al., 2006, Relative and absolute quantification of postsynaptic density proteome isolated from rat forebrain and cerebellum. Mol Cell Proteomics 5:1158-1170.

Chi A et al., 2007, Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. Proc Natl Acad Sci USA 104:2193-2198.

Covey TR, Devanand P, 2002, Nanospray Electrospray Ionization Development: LC/MS, CE/MS Application. Practical Spectroscopy Series, Volume 32: Applied Electrospray Mass Spectrometry.

Dole et al., 1970, Molecular beans from macroions. Journal of Chemical Physics 52:4977-4986.

Duranti M et al., 1981, The seed globulins of *Lupinus albus*. Phytochem 20:2071-2075.

Duranti M et al., 1994a, The unusual extracellular localization of conglutin γ in germinating *Lupinus albus* seeds rules out its role as storage protein. J Plant Physiol 143:711-716.

Duranti M et al., 1994b, Heat-induced synthesis and tunicamycin secretion of putative storage glycoprotein conglutin γ from mature lupin seeds. Eur. J. Bioche, 222, 387-393.

Duranti M et al., 1995a, The saccaride chain of lupin seed conglutin-γ is not responsible for the protection of native protein from degradation by tripsin, but facilitates the refolding of the acid-treated protein to resistant conformation. Eur. J. Biochem 230:886-891.

Duranti M et al., 1995b, Lectin-like activity of lupin seed conglutin γ, a glycoprotein previously referred to as a storage protein. J Exp Bot 46.

Duranti M, Gius C, 1997, Legume seeds: protein content and nutritional value. Field Crop Research 53:31-45.

FDA. Soy protein and coronary heart disease 1999. Federal register 64:57699-57733.

Fenn JB et al., 1989, Electrospray ionization for mass spectrometry of large biomolecules. Science 246:64-71.

Ferreira RB et al., 2003, The structure of *Lupinus* seed storage proteins. Plant Biol 4:139-150.

Fortier M et al., 2005, Integrate microfluidic device for mass spectrometry-based proteomicsand its application to biomarker discovery programs. Anal Chem 77:1631-1641.

Foster LJ et al., 2003, Unbiased quantitative proteomics of lipid rafts reveals high specificity for signalling factors. Proc Natl Acad Sci USA, 100:5813-5818.

Frank AM et al., 2008, Interpreting top-down mass spectra using spectral alignment. Anal Chem 80:2499-2505.

Gartner CA et al., 2007, Catch-and-release reagents for broadscale quantitative proteomics analyses. Journal of Proteome Research 6:1482 1491.

Gayler KR et al., 1990, Biosynthesis, cDNA and amino acid sequences of a precursor of conglutin delta, a sulphur-rich protein from *Lupinus angustifolius*. Plant Mol Biol 15:879-893.

Ge Y et al., 2002, Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. J Am Chem Soc 124:672-678.

Gerber SA et al., 2003, Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci USA. 100:6940-6945.

Ghaemmaghami S et al, 2003, Global analysis of protein expression in yeast. Nature 425:737-741.

Gibbs PEM et al., 1989, Regulation of gene expression during plant embryogenesis. Cell 56:149-160.

Goodlett DR et al., 2001, Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. Rapid Commun Mass Spectrom 15:1214-1221.

Gygi SP et al., 1999, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17:994-999.

Hall RS et al., 2005, Lupin kernel fibre-enriched foods beneficially modify serum lipids in men. Eur J Clin Nutr 59:325-333.

Hardouin J et al., 2006, Usefulness of an integrate microfluidic device (HPLC-Chip-MS) to enhance confidence in protein identification by proteomics. Rapid Commun. Mass Spectrom 20:3236-3244.

Hirano H et al., 1987, Characterization of proteins released from legume seeds in hot water. Phytochemistry 31:731-735.

Holden L et al., 2008, Characterization of IgE binding to lupin, peanut and almond with sera from lupin-allergic patients. Int Arch Allergy Immunol 146:267-276.

Hove EL, 1974, Composition and protein quality of sweet lupin seed. J Sci Food Agric 25:851-859.

Hu Q et al., 2005, The Orbitrap: a new mass spectrometer. J Mass Spectrom 40:430-443.

Hwang SI et al., 2007, Direct cancer tissue proteomics: a method to identify candidate cancer biomarkers from formalin-fixed paraffin-embedded archival tissues. Oncogene 26:65-76.

Ibarrola N et al., 2003, A proteomic approach for quantitation of phosphorylation using stable isotope labeling in cell culture. Anal Chem 75:6043-6049.

Ippel JH et al., 2004, In vivo uniform (15)N-isotope labelling of plants: using the greenhouse for structural proteomics. Proteomics 4:226-234.

Ishihama Y et al., 2005a, Quantitative mouse brain proteomics using culture-derived isotope tags as internal standards. Nat Biotechnol 23:617-621.

Ishihama Y et al., 2005b, Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics 4:1265-1272.

Kazmi S, Krull I, 2001, Proteomics and current state of protein separation science. Pharma Genomics 15-29.

Kinter M, Sherman EN, 2000, Protein sequencing and identification using tandem mass spectrometry. Wiley-Interscience.

Kirkpatrick DS et al., 2005, The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. Methods 35:265-273.

Krijgsveld J et al., 2003, Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. Nat Biotechnol 21:927-931.

Lasonder E et al., 2002, Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. Nature 419:537-542.

Liebler DC, 2002, Introduction to proteomics: tools for the new biology. Humana Press.

Liu H et al., 2004, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem 76:4193-4201.

Macek B et al., 2006, Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. Mol Cell Proteomics 5:949-958.

Mallick P et al., 2007, Computational prediction of proteotypic peptides for quantitative proteomics. Nat Biotechnol 25:125-131.

Magni F, Kienzle MG, 2002, Proteome Analysis. A new challenge for mass spectrometry. RICHMAC Magazine 84:1-5.

Magni C et al., 2004, Conglutin gamma, a lupin seed protein, binds insulin in vitro and reduces plasma glucose levels of hyperglycemic rats. J Nutr Biochem 15:646-650.

Magni C et al., 2005, Two-dimensional electrophoresis and western-blotting analyses with anti Ara h 3 basic subunit IgG evidence the crossreacting polypeptides of *Arachis hypogaea, Glycine max*, and *Lupinus albus* seed proteomes. J Agric Food Chem 53:2275-2281.

Makarov A et al., 2006, Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. J Am Soc Mass Spectrom 17:977-982.

Marchesi M et al., 2008, Hypolipidaemic and anti-atherosclerotic effects of lupin proteins in a rabbit model. Br J Nutr 4:1-4.

Martins JM et al., 2005, Cholesterol-lowering effects of dietary blue lupin (*Lupinus angustifolius* L.) in intact and ileorectal anastomosed pigs. J Lipid Res 46:1539-1547.

McLafferty et al., 2007, Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. FEBS J 274: 6256-6268.

Melo TS et al., 1994, The seed storage proteins from *Lupinus albus*. Phytochemistry 37:641-648.

Mikesh LM et al., 2006, The utility of ETD mass spectrometry in proteomic analysis. Biochim Biophys Acta 1764:1811-1822.

Mirgorodskaya OA et al., 2000, Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using (18)O labeled internal standards. Rapid Commun Mass Spectrom 14:1226-1232.

Miyagi M, Rao KC, 2007, Proteolytic 18O-labeling strategies for quantitative proteomics. Mass Spectrom Rev 26:121-136.

Olsen JV et al., 2005, Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol. Cell Proteomics 4:2010-2021.

Ong SE et al., 2002, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1:376-86.

Peng J, Gygi SP, 2001, Proteomics: the move to mixture. J Mass Spectrom 36:1083-1091.

Perry RH et al., 2008, Orbitrap mass spectrometry: instrumentation, ion motion and applications. Mass Spectrom Rev 27:661-699.

Pilvi TK et al., 2006, Lupin protein attenuates the development of hypertension and normalises the vascular function of NaCl-loaded Goto-Kakizaki rats. J Physiol Pharmacol 57:167-176.

Pratt JM et al., 2006, Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. Nat Protocols 1:1029-1043.

Ramos PCR et al., 1997, Accumulation of a lectin-like breakdown product of β conglutin catabolism in cotyledons of germinating *Lupinus albus*. Planta 203:26-34.

Rao KC et al., 2005, Peptidyl-Lys metalloendopeptidase-catalyzed 18O labeling for comparative proteomics: application to cytokine/lipolysaccharide-treated human retinal pigment epithelium cell line. Mol Cell Proteomics 4:1550–1557.

Rappsilber J et al., 2002, Large-scale proteomic analysis of the human spliceosome. Genome Res 12:1231-1245.

Restani P et al., 1981, Subunit composition of the seed globulin from seeds of *L. albus*, Phytochemistry 20:2077-2083.

Rezaul K et al., 2008, Direct tissue proteomics in human diseases: potential applications to melanoma research. Expert Rev Proteomics 5:405-412.

Rivers J et al., 2007, Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. Mol Cell Proteomics 6:1416-1427.

Ross P.L., et al., 2004, Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3:1154-1169.

Roy SM et al., 2004, Differential expression profiling of serum proteins and metabolites for biomarker discovery. Int J Mass Spectrom 238:163-171.

Sanders SL et al., 2002, Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. Mol Cell Biol 22:4723-4738.

Schetsrieder M, Baeuerlein R, 2009, Proteome research in food science. Chem Soc Rew 38:2600-2608.

Schmidt A et al., 2003, Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI. J Am Soc Mass Spectrom 14:492-500.

Schmidt A et al., 2005, A novel strategy for quantitative proteomics using isotope coded protein labels. Proteomics 5:826.

Semino GA et al., 1985, Effect of bound carbohydrates of the action of tripsin on lupin seed glycoproteins. J Agr. Food Chem 33:196-199.

Shen Y et al., 2002, High efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. Anal Chem 74:4235-4249.

Shen Y et al., 2005, Characterization of the human blood plasma proteome. Proteomics 5:4034-4045.

Silva JC et al., 2005, Quantitative proteomic analysis by accurate mass retention time pairs. Anal Chem 77:2187-2200.

Sirtori CR et al., 1998, Reduction of serum cholesterol by soy proteins: clinical experience and potential molecular mechanisms. Nutr Metab cardiovasc Dis 8:334-340.

Sirtori CR et al., 2004, Proteins of white lupin seed, a naturally isoflavone-poor legume, reduce colesterolemia in rats and increase LDL receptor activity in HepG2 cells. J Nutr 134:18-23.

Sirtori CR et al., 2009, Nutritional and nutraceutical approaches to dyslipidemia and atherosclerosis prevention: focus on dietary proteins. Atherosclerosis 203:8-17.

Staes A et al., 2004, Global differential non-gel proteomics by quantitative and stable labeling of tryptic peptides with oxygen-18. J Proteome Res 3:786-791.

Stemmann O et al., 2001, Dual inhibition of sister chromatid separation at metaphase. Cell 107:715-726.

Syka JE et al., 2004, Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. J Proteome Res 3:621-626.

Udeshi ND et al., 2008, Methods for analyzing peptides and proteins on a chromatographic timescale by electron-transfer dissociation mass spectrometry. Nat Protoc 3:1709-1717.

Vollmer M et al., 2005, Agilent Technologies Application Note 5989-3538EN.

Wait R et al., 2005, Analysis of *Lupinus albus* storage proteins by two-dimensional electrophoresis and mass spectrometry. J Agric Food Chem 53: 4599-4606.

Wang W et al., 2003, Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal Chem 75:4818-4826.

Wei H et al., 2006, Identification and quantification of neuropeptides in brain tissue by capillary liquid chromatography coupled off-line to MALDI-TOF and MALDI-TOF/TOF-MS. Anal Chem 78:4342-4351.

Wiesner J et al., 2008, Application of electron transfer dissociation (ETD) for the analysis of post translational modifications. Proteomics 8:4466-4483.

Wu CC, MacCoss MJ, 2002, Shotgun proteomics: tools for the analysis of complex biological systems. Curr Opin Mol Ther 4:242-250.

Wu CC et al., 2004, Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. Anal Chem 76:4951-4959.

Yao X et al., 2001, Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. Anal Chem 73:2836-2842.

Yates JR et al., 2006, Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. Anal Chem 78:493-500.

Yates JR et al., 2009, Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. Annual Review of Biomedical Engineering 11:49-79.

2. Study 1:

A label-free internal standard method for the differential analysis of bioactive lupin proteins using nano HPLC-Chip coupled with Ion Trap mass spectrometry.

2.1 Aim of study 1

The aim of the present investigation was to develop an HPLC-Chip-MS/MS label-free method based on protein area for the simultaneous qualitative characterization and relative quantification of target storage proteins in the protein extracts of *Lupinus albus* seeds of different cultivars (cv. Adam, Arés, Lucky, and Multitalia). The main feature of the method is the used of an exogenous protein, the Bovine Serum Albumin (BSA), as internal standard for the normalization and the development of two different alghoritm for the differential analysis of a mature protein, i.e the γ-conglutin, and a whole class of protein isoform belonging to the lupin vicilins.

2.2 Flow scheme of study 1

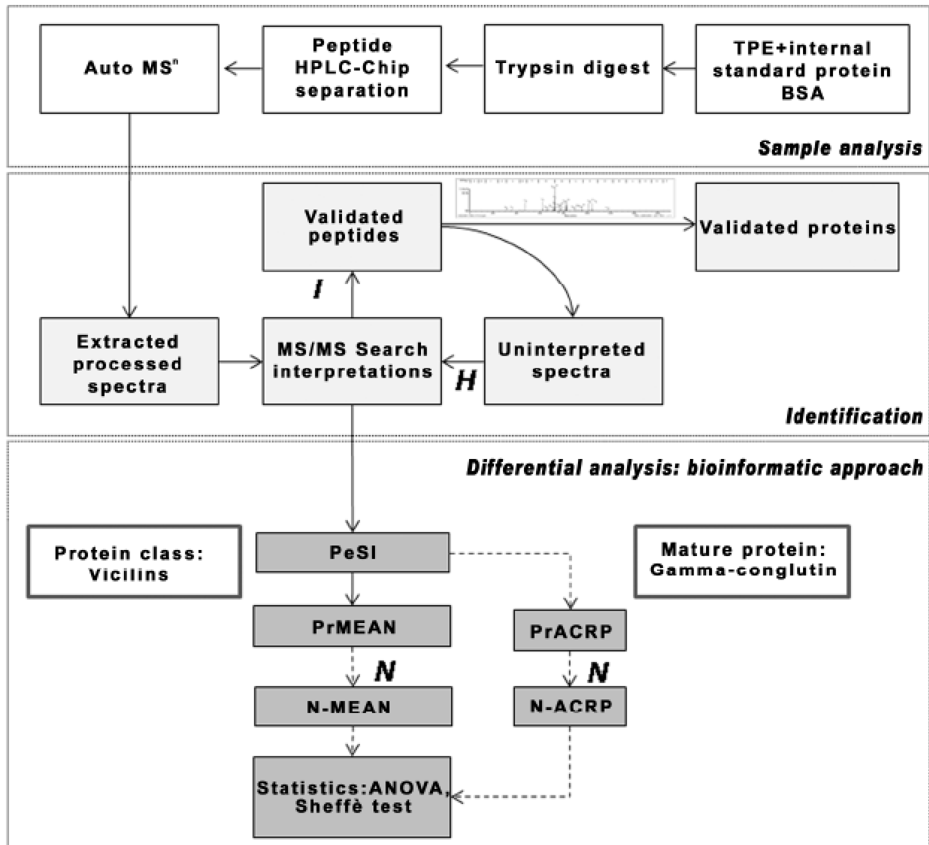


Figure 2.1: Flow scheme for label-free differential analysis of proteins used in study 1: samples to be compared (in principle, an unlimited number of proteolytic digests from un-fractionated total protein extracts) are analyzed by HPLC Chip-MS/MS (three replicates per sample). The raw MS/MS data files are processed using the Spectrum Mill MS Proteomics Workbench. The software extracts the best experimental fragmentation spectra from raw MS/MS data, preprocesses, and searches them in identity mode (I) against the theoretical spectra in a specific database (NCBInr). Each identified and validated peptide is then attributed to a protein sequence which can be validated. The non-validated spectra are searched in homology mode (H) against the validated protein sequences for detecting possible modified and substituted peptides. Spectrum Mill calculates the PeSI from peptide EICs, and the PrMEAN for each identified protein. PeSIs are used for calculating PrACRP. The last step is a normalization procedure (*N*) to derive relative algorithm N-MEAN and N-ACRP.

2.3 Materials and methods

Mature dry seeds of *Lupinus albus* cv. Adam, Arés, Lucky and Multitalia were kindly provided by Dr. Paolo Annicchiarico (CRAISCF, Lodi, Italy). The seed globulins were extracted by the following experimental procedure: defatted lupin flour was extracted with 100 mM Tris-HCl/0.5 M NaCl buffer (pH 8.2) for 2 h at room temperature, with gentle stirring. The solid residue was eliminated by centrifugation at 10 000 rpm, for 20 min at 4° C and the supernatant was dialyzed against 30 mM Tris-HCl buffer (pH 8.2) for 24 h at 47C.
The protein content was assessed according to Bradford [Bradford, 1976]. The protein extracts from the different cultivars of *L. albus* were digested in the presence of a constant amount of BSA (1 ng BSA : 10 ng protein extract), for the simultaneous qualitative and quantitative analysis of lupin proteins (Figure 2.1).
Each mixture was denaturated with 6 M urea and reduced by adding 200 mM 1,4-DTT (in the ratio of 50 mol DTT: 1 mol Cys), reacting at room temperature for 1 h. The sample was then alkylated with 200 mM iodoacetamide (IAM, in the ratio of 200 mol IAM: 1 mol Cys), and the mixture was kept at room temperature in the dark for 1 h. The excess of IAM was neutralized by adding an equimolar amount of DTT. In order to reduce the urea concentration to 0.6 M, the solution was diluted to 1 mL with buffer and then digested with sequencing grade trypsin (0.5 mg/mL) in the ratio 1 : 50 enzyme/protein w/w at 37°C over night. A 2 uL aliquot of each tryptic digest (in total 110 ng proteins = 100 ng protein extract + 10 ng BSA) was analyzed by HPLC-Chip-IT MS/MS in three replicates. Samples were injected onto a LC/MS system consisting of a 1200 Series liquid chromatograph, an HPLC-Chip Cube MS interface, and SL IT mass spectrometer (all Agilent Technologies, Palo Alto, CA, USA). The chromatographic chip incorporated a 40 nL enrichment column, a 43 mm x 75 mm analytical column packed with Zorbax 300SB-C18 5 mm particles and a nanospray needle. Peptides were loaded onto the enrichment column before the analytical separation: the capillary pump delivered an isocratic 100 % C solvent phase (99 % water, 1 % ACN, and 0.1 % formic acid) at 4 uL/min. Ssolvent A was: 95 % water, 5 % ACN, 0.1 % formic acid; solvent B was: 95 % ACN, 5 % water, 0.1 % formic acid. The nano-pump gradient program was as follows: 3 % solvent B (0 min), 50 % solvent B (0-50 min), 80 % solvent B (50-55 min), 80 % solvent B for 5 min and back to 3 % in 10 min at 0.3 uL/min. To ensure optimized nano-flow and fast gradient response, both capillary and nano-pump flows were controlled using nano-flow sensors and active splitters. The drying gas temperature was 300°C, the flow was 3 L/min (nitrogen), data acquisition occurred in positive ionization mode. Capillary voltage was -1850 V with endplate offset of - 2500 V. The recorded mass range was 300-2200 m/z, target mass 700 m/z, average of 2 spectra, ICC target 30 000, maximum accumulation time was 150 ms. The MS/MS analyses were performed in Auto MSn mode: the fragmentation amplitude was set to 1 V, the number of MS/MS stages was 2, and the number of precursor ions selected for MS/MS during each scan was 2, doubly charged ions were preferred.

The mass peak intensity threshold to trigger the Auto MSn experiments was the same for all samples. Figure 2.1 shows the experimental strategy used in this work to derive simultaneously qualitative and quantitative information. The whole National Center for Biotechnology Information non redundant (NCBInr) database was searched using the Agilent Spectrum Mill software (Rev A.03.03). The Spectrum Mill Data Extractor program prepares MS/MS data files for processing: it extracts high-quality experimental fragmentation spectra from raw MS/MS data files and rejects spectra that are too noisy or do not represent peptides. The MS/MS raw file data extractor extracts and merges spectra with the same precursor ion within ± 1.4 *m/z* and within a time frame of ± 15 s. MS/MS spectra are preprocessed: the extractor attributes precursor charge (up to 7), centroides the MS/MS spectra, calculates different spectral features, filters MS/MS spectra by quality and calculates extracted ion chromatograms (EICs) for the MS precursor scans. Precursor ions need to have a minimum S/N value of 25 and the C12 peaks are determined by the Data Extractor. The preprocessed spectra are used by Spectrum Mill to interrogate the NCBInr protein database in the MS/MS search; NCBInr is searched for tryptic peptides with a mass tolerance of ± 2.5 Da for precursor ions and a tolerance of ± 0.7 Da for fragment ions. One missed cleavage is allowed. In the first stage of MS/MS search, the identity mode is used to look for sequences that are identical to the peptide used to generate the MS/MS spectrum (unmodified peptide). For each identified peptide Spectrum Mill calculates the peptide score, which is based on a point system: points are added for each peak assigned to an allowed fragment ion type for a candidate peptide sequence, different types of fragment ions are worth different points. Points are subtracted for each unassigned peak and the penalty value is based on the unassigned peak height. Besides the scores, the percentage Scored Peak Intensity (SPI %) is an important parameter which indicates the percentage of total ion intensity of the MS/MS spectrum assigned to product ions that are indicative of peptides. The protein score is obtained by adding the scores of single peptides. The Spectrum Mill software is able to perform a reversed database search. For a reversed database search, the software reverses only the internal portion of the peptide sequences in the database. All of these internally reversed sequences from the database are compared to the MS/MS experimental spectrum and the one that returns the highest score is saved as the reversed database hit. The reversed database hit is not always the reverse of the peptide that matched in the forward search, because a different reversed hit may score higher. The forward–reversed score is the difference between scores for top hits from forward and reversed database searches. Besides the forward–reversed score, Spectrum Mill calculates the rank 1–2 score which is the difference between the scores of the top and second highest scoring database hit. Spectrum Mill default values used to validate the peptides at the different charge states are: peptide score > 11, SPI > 60 %, forward–reversed score > 2, rank1–rank2 score > 2 for the charge state 2+; peptide score > 13, SPI > 70 %, forward–reversed score > 2, rank1–rank2 score > 2 for the charge state 1+; peptide score > 13, SPI > 70 %, forward–reversed score > 2, rank1–rank2 score > 2 for the charge state 3+. The peptide score and SPI % thresholds assure a high quality of the match between experimental and theoretical fragmentation spectra and, at the same time, the forward–reversed score and rank1–rank2 score thresholds help to rule out false positives.

The validated protein sequences show protein scores higher than 20 in all analyses and the resulting validated protein sequences are used as hits in the second MS/MS search step that is performed in the homology mode to search the un-validated MS/MS spectra. The homology mode looks both for the main variable modification among a limited list of modified amino acids (modified peptide: acetyl K, oxidized M, pyroglutamic acid on N-terminal Q, deaminated N, phosphorilated S, T, Y), and for matches that are consistent with a single aminoacid substitution (substituted peptides). The MS/MS spectra of modified and substituted peptides are validated with the same thresholds reported above. The homology mode search was useful for increase in the coverage of validated proteins. Consequently, the homology MS/MS search resulted to be an essential step in the flowscheme of the method to assure a satisfactory qualitative and quantitative analysis.

2.4 Results

2.4.1 Qualitative analysis of the storage proteins

The main classes of lupin seed storage proteins are the following: 7S acidic globulins (β-conglutin or vicilins), a 7S basic globulin (γ-conglutin), 11S globulins (α-conglutin or legumins), and a 2S globulin (δ-conglutin). Two papers have reported the 2-DE profiles of these proteins: the former is based on the cultivar Arés, while the latter on the cultivar Multitalia [Magni et al., 2007]. Whereas γ-conglutin is a homogeneous protein, composed by a heavy and a light chain linked by disulfide bonds, the vicilins and the legumins have a multigenic origin and appear as complex mixtures of polypeptides with different molecular weights and pIs. A single HPLC-Chip-MS/MS analysis of the protein extract tryptic digest permitted to identify all the lupin seed storage proteins with very satisfactory coverages: major and minor proteins were simultaneously identified without any previous fractionation of the extracts. The lupin protein sequences successfully identified in the protein database NCBInr were the following: for the class of vicilins, the β-conglutin precursor (NCBInr accession no. 46451223) and the vicilin-like protein (NCBInr accession no.89994190); for γ-conglutin, the sequence NCBInr accession no. 11191819; for the class of legumins, the legumin-like protein (NCBInr accession no. 85361412); for δ-conglutin, the δ-conglutin seed storage protein precursor (NCBInr accession no. 80221495). In the case of the vicilins, the HPLC-Chip-MS/MS analysis enabled the identification of both the β-conglutin precursor with a coverage ranging from 32 to 49 % and the vicilin-like protein with a coverage ranging from 33 to 42 % related to the four cultivars. In spite of their high sequence homology, estimated by using the BLAST-P program as 78 % identity and 80 % positivity, it was possible to identify both common and specific peptides of these sequences. The identified specific peptides of the β-conglutin precursor were indexed as 34, 36, 43, 44, 47, 54, 59, 60 in Table 2.1, whereas those of the vicilin-like protein as 38, 39, 50, 55, 57, 61, 71.

Table 2.1: Qualitative characterization of all the lupin storage proteins. For each class are reported the identified sequences and their "cumulative covarages" (sequence coverage obtained by merging the qualitative analyses of the proteins in the four cultivars). For each protein sequence are reported all the identified peptides (unmodified, modified and substituted peptides). Each peptide is characterized by the index number, the peptide sequence, the matched MH+ , the pI, the state charge (z) and their start position along the protein sequence. For the modified and substituted peptides the variable site is reported too. a Single aminoacid substitutions and variable modifications; b Position of the start aminoacid in protein sequence

A) Vicilin:

beta-conglutin precursor (NCBInr accession number 46451223): cumulative coverage 52 %

vicilin-like protein (NCBInr accession number 89994190): cumulative coverage 49 %

Index	Peptide sequence	Matched MH^+(Da)	p*I*	Protein identification	z	Variable site[a]	Start AA position[b]
1.	(R)TNRLENLQNYR(I)	1420.729	8.41	beta conglutin precursor, vicilin-like protein	3	-	143
2.	(R)IVEFQSKPNTLILPK(H)	1727.01	8.59	beta conglutin precursor, vicilin-like protein	3	-	154
3.	(R)RVEFQSKPNTLILPK(H)	1727.01	9.99	beta conglutin precursor, vicilin-like protein	3	I154R	154
4.	(K)HSDADYVLVVLNGR(A)	1557.802	5.21	beta conglutin precursor, vicilin-like protein	2, 3	-	169
5.	(K)HVDADYVLVVLNGR(A)	1557.802	5.21	beta conglutin precursor, vicilin-like protein	3	S170V	169
6.	(K)HSDAQYVLVVLNGR(A)	1557.802	6.74	beta conglutin precursor, vicilin-like protein	3	D173Q	169
7.	(K)HSQADYVLVVLNGR(A)	1557.802	6.74	beta conglutin precursor, vicilin-like protein	3	D171Q	169
8.	(K)HSDAKYVLVVLNGR(A)	1557.802	8.60	beta conglutin precursor, vicilin-like protein	3	D173K	169

9.	(K)HSKADYVLVVLNGR(A)	1557.802	8.60	beta conglutin precursor, vicilin-like protein	3	D171K	169
10.	(K)HMDADYVLVVLNGR(A)	1557.802	5.21	beta conglutin precursor, vicilin-like protein	3	S170M	169
11.	(K)HSDNDYVLVVLNGR(A)	1557.802	5.21	beta conglutin precursor, vicilin-like protein	3	A172N	169
12.	(R)ATITIVNPDRR(Q)	1255.712	9.64	beta conglutin precursor, vicilin-like protein	3	-	183
13.	(R)ATITIVNPDR(R)	1099.611	5.88	beta conglutin precursor, vicilin-like protein	2	-	183
14.	(R)QAYNLEYGDALR(I)	1412.68	4.37	beta conglutin precursor, vicilin-like protein	2	-	194
15.	(R)QAYPLEYGDALR(I)	1412.68	4.37	beta conglutin precursor, vicilin-like protein	2	N197P	194
16.	(R)qAYNLEYGDALR(I)	1412.68	4.37	beta conglutin precursor, vicilin-like protein	2	Q194q	194
17.	(R)QAYNLEYGDALRIPAGSTSYILNPDDNQK(L)	3226.565	4.23	beta conglutin precursor, vicilin-like protein	3	-	194
18.	(R)QAYPLEYGDALRIPAGSTSYILNPDDNQK(L)	3226.565	4.23	beta conglutin precursor, vicilin-like protein	3	N197P	194
19.	(R)QAYNLEYGDALRIPAGSTSYINNPDDNQK(L)	3226.565	4.23	beta conglutin precursor, vicilin-like protein	3	L215N	194
20.	(R)IPAGSTSYILNPDDNQK(L)	1832.902	4.21	beta conglutin precursor, vicilin-like protein	2, 3	-	206
21.	(R)IPAGSTSYILNPDDNQKLR(V)	2102.088	5.96	beta conglutin precursor, vicilin-like protein	3	-	206
22.	(R)VVKLDIPINNPGYFYDFYPSSTK(D)	2633.365	5.93	beta conglutin precursor, vicilin-like protein	3	A229D	225
23.	(R)VVKLNIPINNPGYFYDFYPSSTK(D)	2633.365	8.35	beta conglutin precursor, vicilin-like protein	3	A229N	225
24.	(R)VVRLAIPINNPGYFYDFYPSSTK(D)	2633.365	8.40	beta conglutin precursor, vicilin-like protein	3	K227R	225
25.	(R)VVKLARPINNPGYFYDFYPSSTK(D)	2633.365	9.40	beta conglutin precursor, vicilin-like protein	3	I230R	225

26.	(R)VVKRAIPINNPGYFYDFYPSSTK(D)	2633.365	9.40	beta conglutin precursor, vicilin-like protein	3	L228R	225
27.	(R)VVKLVIPINNPGYFYDFYPSSTK(D)	2633.365	8.35	beta conglutin precursor, vicilin-like protein	3	A229V	225
28.	(K)LAIPINNPGYFYDFYPSSTK(D)	2307.133	5.83	beta conglutin precursor, vicilin-like protein	2, 3	-	228
29.	(K)DQQSYFSGFSR(N)	1321.581	5.83	beta conglutin precursor, vicilin-like protein	2	-	248
30.	(R)NTLEATFNTR(Y)	1166.58	6.00	beta conglutin precursor, vicilin-like protein	2	-	259
31.	(R)NTLEATFNTRYEEIQR(I)	1984.972	4.79	beta conglutin precursor, vicilin-like protein	3	-	259
32.	(R)IILGNEDEQEYEEQRR(G)	2020.957	4.14	beta conglutin precursor, vicilin-like protein	3	-	275
33.	(R)IILGNEDEQEYEEQR(R)	1864.856	3.83	beta conglutin precursor, vicilin-like protein	2, 3	-	275
34.	(R)RGQEQSDQDEGVIVISSK(K)	1987.009	4.32	beta-conglutin precursor	3	V305S	290
35.	(K)YGNFYEITPDR(N)	1374.632	4.37	beta conglutin precursor, vicilin-like protein	2	-	345
36.	(K)INEGALLLPHYNSK(A)	1568.843	6.75	beta-conglutin precursor	2, 3	-	371
37.	(K)RNEGALLLPHYNSK(A)	1568.843	8.60	beta-conglutin precursor	2	I371R	371
38.	(K)AIFIVVVGEGNGK(Y)	1302.742	6.05	vicilin-like protein	2	-	385
39.	(K)AIFIVVVGEGNGKYELVGIR(D)	2133.207	6.19	vicilin-like protein	3	-	385
40.	(K)AIFIVVVGEGnGKYELVGIR(D)	2133.207	6.19	vicilin-like protein	3	N395n	385
41.	(K)AIFIVVVGEGDGKYELVGIR(D)	2133.207	6.19	vicilin-like protein	3	N395D	385
42.	(K)AIFIVVVGEGNGEYELVGIR(D)	2133.207	4.25	vicilin-like protein	3	K397E	385

43.	(K)AIYVVVVDEGEGNYELVGIR(D)	2194.139	4.00	beta-conglutin precursor	3	-	385
44.	(K)AIYVVVVDEGEGNYELVGIRDQQR(Q)	2721.384	4.18	beta-conglutin precursor	3	-	385
45.	(K)AIYTVVVDEGEGNYELVGIRDQQR(Q)	2721.384	4.18	beta-conglutin precursor	3	V388T	385
46.	(K)ADYVVVVDEGEGNYELVGIRDQQR(Q)	2721.384	4.02	beta-conglutin precursor	3	I386D	385
47.	(R)LSEGDIFVIPAGYPISINASSNLR(L)	2533.33	4.00	beta-conglutin precursor	3	-	427
48.	(R)LSEGDIFVIPAGYPTSINASSNLR(L)	2533.33	4.37	beta-conglutin precursor	3	I441T	427
49.	(R)LSEGDIFVIPAGYPISTNASSNLR(L)	2533.33	4.37	beta-conglutin precursor	3	I443T	427
50.	(R)LSEGDIFVIPAGYPISVNASSNLR(L)	2519.314	4.37	vicilin-like protein	3	-	427
51.	(R)LSEGDIFVIIAGYPISVNASSNLR(L)	2519.314	4.37	vicilin-like protein	3	P436I	427
52.	(R)LSEGDIFVILAGYPISVNASSNLR(L)	2519.314	4.37	vicilin-like protein	3	P436L	427
53.	(R)LSEGDIFVIPSGYPISVNASSNLR(L)	2519.314	4.37	vicilin-like protein	3	A437S	427
54.	(R)LLGFGINADENQR(N)	1446.734	4.37	beta-conglutin precursor	2, 3	-	451
55.	(R)LLGFGINAYENQR(N)	1494.77	6.00	vicilin-like protein	2, 3	-	451
56.	(R)RLGFGINAYENQR(N)	1494.77	8.75	vicilin-like protein	2	L451R	451
57.	(R)NFLAGSEDNVIR(Q)	1334.67	4.37	vicilin-like protein	2	-	464
58.	(R)NFLAGFEDNVIR(Q)	1334.67	4.37	vicilin-like protein	2	S469F	464
59.	(R)NFLAGSKDNVIR(Q)	1333.722	8.75	beta-conglutin precursor	2	-	464

60.	(R)AVNELTFPGSAEDIER(L)	1747.85	4.00	beta-conglutin precursor	2	-	480
61.	(K)ELTFPGSAEDIER(L)	1463.701	4.00	vicilin-like protein	2	-	483
62.	(K)ERTFPGSAEDIER(L)	1463.701	4.41	vicilin-like protein	2	L484R	483
63.	(K)ELTFPGSADDIER(L)	1463.701	3.91	vicilin-like protein	2	E491D	483
64.	(K)ELTFPGSAEDVER(L)	1463.701	4.00	vicilin-like protein	2	I493V	483
65.	(K)ELTFPGSAEDIDR(L)	1463.701	3.91	vicilin-like protein	2	E494D	483
66.	(K)ELTFLGSAEDIER(L)	1463.701	4.00	vicilin-like protein	2	P487L	483
67.	(K)ELTFPGSSEDIER(L)	1463.701	4.00	vicilin-like protein	2	A490S	483
68.	(K)ELTFIGSAEDIER(L)	1463.701	4.00	vicilin-like protein	2	P487I	483
69.	(K)ELTFPGSAETIER(L)	1463.701	4.25	vicilin-like protein	2	D492T	483
70.	(K)ELTFPGSGEDIER(L)	1463.701	4.00	vicilin-like protein	2	A490G	483
71.	(R)LIKnQQQSYFANALPQQQQQSEK(E)	2719.38	8.50	vicilin-like protein	3	N499n	496

B) γ-conglutin

conglutin-gamma (NCBInr accession number 11191819): cumulative coverage 33%

Index	Peptide sequence	Matched MH^+(Da)	p*I*	Protein Name	z	Variable site[a]	Start AA position[b]
72.	(K)RTPLMQVPVLLDLNGK(H)	1794.031	8.75	conglutin	3	-	68
73.	(R)TPLMQVPVLLDLNGK(H)	1637.93	5.50	conglutin gamma	3	-	69

74.	(K)IPQFLFSCAPTFLTQK(G)	1897.988	8.22	conglutin gamma	3, 2	-	172
75.	(K)QGEYFIQVSAIR(V)	1410.738	6.00	conglutin gamma	2, 3	-	268
76.	(K)qGEYFIQVSAIR(V)	1410.738	6.00	conglutin gamma	2	Q268q	268
77.	(R)HSIFEVFTQVFANNVPK(Q)	1977.023	6.75	conglutin gamma	3	-	324
78.	(K)AVGPFGLCYDTK(K)	1327.635	5.87	conglutin gamma	2	-	346
79.	(K)ISGGVPSVDLIMDK(S)	1430.756	4.21	conglutin gamma	2	-	359
80.	(K)ISGGVPSVDLIMDKSDVVWR(I)	2173.132	4.54	conglutin gamma	3	-	359
81.	(R)ISGENLMVQAQDGVSCLGFVDGGVHTR(A)	2846.356	4.54	conglutin gamma	3	-	379
82.	(R)AGIALGTHQLEENLVVFDLAR(S)	2266.219	4.65	conglutin gamma	3	-	406
83.	(K)SCSNLFDLNNP(-)	1280.558	3.80	conglutin gamma	2	-	442

C) Delta-conglutin

conglutin-delta seed storage protein precursor (NCBInr accession number 80221495)**:** cumulative coverage 53%

Index	Peptide sequence	Matched MH^{+}(Da)	p*I*	Protein Name	z	Variable site[a]	Start AA position[b]
84.	(K)SQLQQVNLNHCENHIIQR(I)	2231.11	6.66	conglutin delta seed storage protein precursor	3	-	32
85.	(K)SQLQQVNLnHCENHIIQR(I)	2231.11	6.66	conglutin delta seed storage protein precursor	3	N40n	32
86.	(K)SQLQQVnLNHCENHIIQR(I)	2231.11	6.66	conglutin delta seed storage protein precursor	3	N38n	32
87.	(R)SSQESEELDQCCEQLNELNSQR(C)	2683.121	3.83	conglutin delta seed storage protein precursor	3	-	75

Index	Peptide sequence	Matched MH+(Da)	pI	Protein Name	z	Variable site[a]	Start AA position[b]
88.	(R)SSQESEELDQCCEQLnELNSQR(C)	2683.121	3.83	conglutin delta seed storage protein precursor	3	N90n	75
89.	(R)ALQQIYENQSEQCQGR(Q)	1951.893	4.53	conglutin delta seed storage protein precursor	3, 2	-	101
90.	(R)ALQQIYENQSEQCAGR(Q)	1951.893	4.53	conglutin delta seed storage protein precursor	3	Q114A	101
91.	(R)QEEQLLEQELENLPR(T)	1867.94	3.98	conglutin delta seed storage protein precursor	3, 2	-	117
92.	(R)QEEQLLEQELEPLPR(T)	1867.94	3.96	conglutin delta seed storage protein precursor	2	N128P	117
93.	(R)qEEQLLEQELENLPR(T)	1867.94	3.98	conglutin delta seed storage protein precursor	2	Q117q	117
94.	(R)QEEQLLEQELENLPRFCGFGPLR(R)	2756.367	4.32	conglutin delta seed storage protein precursor	3	T132F	117

D) Legumin:

legumin-like protein (NCBInr accession number 85361412): cumulative coverage 18%

Index	Peptide sequence	Matched MH+(Da)	pI	Protein Name	z	Variable site[a]	Start AA position[b]
95.	(R)LNALEPDNTVQSEAGTIETWNPK(N)	2527.231	4.00	legumin-like protein	3	-	41
96.	(R)RPFYTNAPQEIYIQQGR(G)	2081.056	8.59	legumin-like protein	3	-	86
97.	(R)RFYLSGNQEQEFLQYQEK(E)	2307.104	4.79	legumin-like protein	3	-	185
98.	(R)FYLSGNQEQEFLQYQEK(E)	2151.003	4.25	legumin-like protein	3	-	186
99.	(K)TLTSIDFPILGWLGLAAEHGSIYK(N)	2602.392	5.29	legumin-like protein	3	-	360
100.	(K)FLVPPPQSQLR(A)	1281.731	9.75	legumin-like protein	2	-	499

Vicilins are oligomers of 150-170 kDa formed by three similar 40-70 kDa subunits with no disulfide linkages. They are highly heterogeneous and their heterogeneity is due to the expression of multigene families whose individual genes are very closely related. cDNA deduced vicilin sequences in databases are not complete yet. In this work, the homology search was performed with the aim of complementing the information that can be obtained by an identity search against the deposited vicilin-like protein and β-conglutin precursor sequences. Several single aminoacid substitutions were successfully identified confirming the presence of high homology vicilin precursors; the MS/MS spectra of substituted peptides were validated with the same thresholds of the unmodified peptide and were all manually interpreted in order to confirm all mutations (Table 2.1). Some mutations have been identified in all four cultivars: S469F, I493V, E494D, D492T. All the other single aminoacid mutations are not common to all cultivars according to the high degree of polymorphism within each species of lupin [Freitas et al., 2007]. Among the considered variable modifications, only the N deamination and the Q N-term pyroglutamination were identified: Q194q, N395n, N499n (Table 2.1). γ - Conglutin was identified with coverages ranging from 25 to 33 % in all four cultivars, a very good result considering that this protein is very resistant to enzymatic digestion. In the γ - conglutin sequence the homology search did not highlight any single aminoacid substitution confirming that this protein is very homogeneous [Duranti et al., 1981] (Table 2.1). The percentage sequence coverage of the legumin-like protein was always small falling between 3 and 13 %, which may perhaps indicate that the correct sequences have not been deposited yet in the database; this is not uncommon for plants, since only the genomes of *Arabidopsis thaliana* [Gallardo et al., 2003], *Medicago truncatula* [Gallardo et al., 2001; Gallardo et al., 2002], *Glycine max* [Hajduch et al., 2005], and *Triticum aestivus* [Majoul et al., 2003] have been fully or extensively characterized. The coverage of δ-conglutin was between 27 and 37 %.

2.4.2 Label-free differential analysis of the seed storage proteins

Spectrum Mill is able to extract EICs of all precursor ions and to use them for peptide quantification. The peptide spectrum intensity (PeSI in Figure 2.1) is the chromatographic peak area of each peptide precursor ion, which is calculated by summing the precursor m/z abundance in the MS scans within a time frame of ± 15 s and within a mass tolerance of ± 1.4 m/z. The parameter “protein mean peptide spectral intensity”(PrMEAN, Figure 2.1) is the mean peak intensity of all peptide precursor ions identified for each protein. This parameter was calculated for each single analysis (three replicates, i.e., X1, X2, and X3), giving PrMEAN1, PrMEAN2, and PrMEAN3. They were then averaged to PrMEAN (Table 2.2A) for each cultivar.

Table 2.2: The most important analytical parameters used in the study 1. X1, X2, X3 represent the replicates of a single cultivar. Common Reproducible Peptides (CRPs) are those peptides identified in all 3 replicates of the same cultivar and in all cultivars. Reproducible Specific Peptides (SRPs) are those peptides identified in all 3 replicates of the same cultivar, but not in all cultivars. Non-Reproducible Peptides (NRPs) are those peptides identified only in some replicates of a single cultivar. In pane A) the Protein Mean Peptide Spectral Intensity (PrMEAN) is the mean peak intensity of all peptide precursor ions identified for each protein. This parameter is calculated for each single analysis (3 replicates, i.e. X1, X2, X3), giving PrMEAN1, PrMEAN2, PrMEAN3, which are then averaged to PrMEAN for each cultivar. In pane B) the Protein Average of Common Reproducible Peptides (PrACRP) is the mean peak intensity of just the CRPs identified for each protein for each single analysis (3 replicates, i.e. X1, X2, X3), giving PrACRP1, PrACRP2, PrACRP3, which are then averaged to PrACRP.

A)

	Sample	
X_1	X_2	X_3
CRP	CRP	CRP
SRP	SRP	SRP
NRP	*NRP*	*NRP*
$PrMEAN_1$	$PrMEAN_2$	$PrMEAN_3$
	PrMEAN	

B)

	Sample	
X_1	X_2	X_3
CRP	CRP	CRP
$PrACRP_1$	$PrACRP_2$	$PrACRP_3$
	PrACRP	

In Table 2.2A, peptides identified for the target proteins in each replicate were classified as "common reproducible peptides" (CRPs), "specific reproducible peptides" (SRPs), and "nonreproducible peptides" (NRPs). CRPs are those peptides identified in all three replicates of the same cultivar and in all cultivars, independently of their RSD %; SRPs are those peptides identified in all three replicates of the same cultivar, but not in all cultivars; whereas NRPs are those peptides identified only in some replicates of singular cultivar. Therefore, another potentially useful parameter was calculated and named "protein average of common reproducible peptides" (PrACRP, Table 2.2B), which is the mean peak intensity of just the CRPs identified for each protein. The aim of this new parameter was to reduce the variability of the data due to the contribution of SRPs and NRPs. Finally, in order to compare the different samples, they were spiked with an exogenous internal standard (BSA) and the two parameters just described were normalized (N in Figure 2.1) by dividing each of them by the corresponding parameters of BSA. Thus, the PrMEANs of the target lupin proteins were divided by the PrMEANs of BSA to give the normalized parameters, "normalized protein mean peptide spectral intensity" (N-MEANs), and the Pr-ACRPs of the target lupin proteins were divided by the PrACRPs of BSA to give the normalized parameters, "normalized protein average of common reproducible peptides" (N-ACRPs) (Figure 2.1).The normalized parameters were submitted to a statistical evaluation through one way ANOVA ($p < 0.05$) and the Sheffè test to verify whether there were any statistically significant differences in the content of the target proteins, i.e., γ-conglutin and vicilins, among the four investigated cultivars.

2.4.2.1 Quantitative analysis of the internal standard BSA

A very critical point in the application of a method based on LC-MS/MS to different samples is the possible presence of matrix effect. Possible approaches to address this problem could be complex clean-up procedures or a reduction of the sample complexity by fractionation prior to LC-MS/MS analysis. Both approaches may impair the reproducibility and modify extensively the composition of the starting materials. Considering all these facts, intensity-based label-free differential analysis is considered reliable only when it involves the use of an internal standard for minimizing the matrix effect, which would lead to a variation in MS response by ion suppression of peptides due to the presence of coeluting components [Tabata et al., 2007]. Normalization procedures may be particularly important for minimizing systematic biases in ion intensities introduced by sample handling, sample concentration, and instrument sensitivity drifts during the course of data acquisition [Higgs et al., 2005]. Protein extracts from four different lupin cultivars were spiked with a fixed amount of highly pure BSA (1 ng BSA: 10 ng protein extract) prior to enzymatic digestion. This ratio was considered ideal among a few investigated ratios, because it did not modify significantly the identification of lupin proteins and enabled the identification of BSA with a satisfactory coverage ranging from 11-33 %. BSA was identified in all samples with four CRPs that were reported in Table 2.3.

Table 2.3: The CRPs (Common Reproducible Peptides) of the internal standard BSA in the four cultivars (*cv.* Adam, Arés, Lucky, Multitalia) are reported. For each peptide the index number, the sequence, the average intensity in the three replicates and the corresponding standard deviation and relative standard deviation percentage (RSD %) are reported. The average of the PrACRPs (Protein Average of Common Reproducible Peptides) and of the PrMEANs (Protein Mean Peptide Spectral Intensities), the standard deviations and relative standard deviation percentages (RSD %) are reported too.

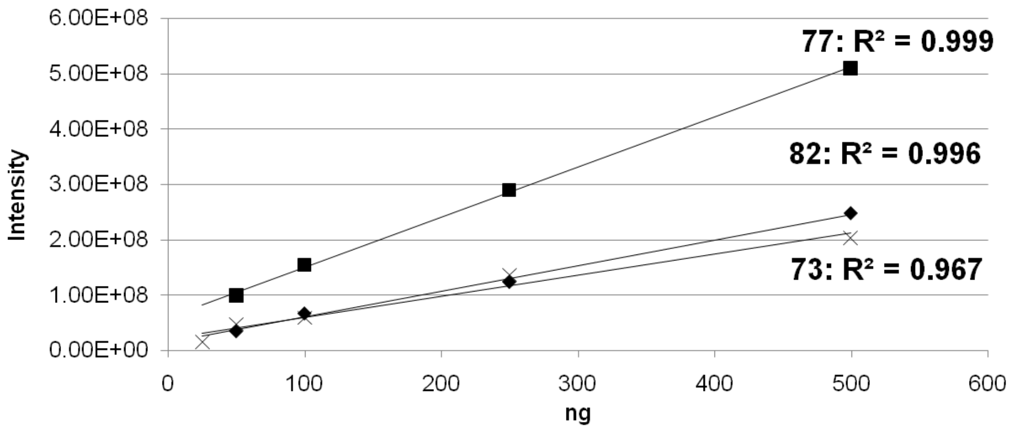
		Matrix											
Index	**CRP BSA**	**Adam**			**Arés**			**Lucky**			**Multitalia**		
		PeSI average	**±SD**	**RSD%**	**PeSI average**	**±SD**	**RSD%**	**PeSI average**	**±SD**	**RSD%**	**PeSI average**	**±SD**	**RSD%**
1.	**DAFLGSFLYEYSR**	6.65E+07	1.18E+07	17.69	6.38E+07	1.81E+06	2.85	5.01E+07	2.95E+06	5.90	4.07E+07	5.36E+06	13.18
2.	**LGEYGFQNALIVR**	6.86E+07	2.95E+07	43.00	3.85E+07	6.67E+06	17.33	5.63E+07	1.03E+07	18.21	3.67E+07	1.32E+07	36.12
3.	**LVNELTEFAK**	1.10E+08	6.08E+06	5.53	6.16E+07	7.50E+06	12.18	9.23E+07	1.35E+07	14.67	1.26E+08	1.71E+07	13.56
4.	**RHPYFYAPELLYYANK**	3.19E+07	2.99E+06	9.36	2.37E+07	5.19E+06	21.88	3.91E+07	6.09E+06	15.58	2.49E+07	6.43E+06	25.84
	PrACRP	6.93E+07	6.64E+06	9.59	4.69E+07	2.76E+06	5.88	5.94E+07	6.55E+06	11.02	5.71E+07	7.07E+06	12.39
	PrMEAN	6.21E+14	8.78E+06	14.14	4.98E+07	5.46E+06	10.96	5.78E+07	5.06E+06	8.75	4.93E+07	4.53E+06	9.19

The intensities of these peptides were used to calculate the PrACRPs. The averages, the SDs and the corresponding RSD % of PrACRPs and PrMEANs are reported in Table 2.3. The RSD % of PrACRPs ranged between 6 and 12 %, whereas that of PrMEANs between 9 and 14 %. PrACRPs and PrMEANs of BSA were used in the differential analyses of γ-conglutin and vicilins to calculate the N-ACRPs and N-MEANs.

2.4.2.2 Quantitative analysis of γ-conglutin

The optimized procedure permitted to detect at least one peptide deriving from γ-conglutin loading only 10 ng of protein extract tryptic digest on the HPLC-Chip. This means that the sensitivity was increased by a factor of 50 with respect to a previous work [Locati et al., 2006], in which the preliminary label-free evaluation on lupin storage proteins was performed through conventional HPLC-ESI-MS/MS shotgun proteomics. In fact, the previous method could detect at least one peptide deriving from γ-conglutin injecting 500 ng of protein extract tryptic digest. Moreover, the use of the HPLC-Chip increased the cumulative coverage of this protein up to 33 % (Table 2.1) whereas in the previous paper the coverage was only 19%. This may be considered a very good result since γ-conglutin is a minor protein in the protein extract, corresponding to less than 5% [Wait et al., 2003] and is rather resistant to protease digestion. The method allowed the identification of γ-conglutin with three CRPs, belonging either to the light chain (82 and 77) or the heavy chain (73). These three peptides were detected in the analyses of all cultivars and in all replicates for the same cultivar (X1, X2, X3) and were, therefore, the best candidates to develop a quantitative method for γ-conglutin. This protein is a mature protein characterized by the same structure in the four cultivars investigated in this work: direct HPLC-Chip MS/MS analysis of the spots of γ-conglutin excised from 2-DE gels of each cultivar permitted the identification of the same peptides (82 and 77) in the spots of the light chain and of the peptide 73 in the spots of the heavy chain. During the development of the method, the linearity of the mass spectrometric response for the intensities of CRPs, PrACRP, and PrMEAN was evaluated injecting different amounts of the unspiked protein extract tryptic digest of the cultivar Arés in the range from 25 to 500 ng. The peptides 82 and 77 showed a linear behavior in the range from 50 to 500 ng tryptic digest loaded on the chip, with correlation coefficients R^2 of 0.996 and 0.999, respectively (Figure 2.2A); whereas the peptide 73 had a R^2 of 0.967 in the range from 25 to 500 ng (Figure 2.2A). Consequently, PrACRP was calculated in the range from 50 to 500 ng, where these CRPs were all detected. Figure 2.2B compares the parameter PrACRP *vs.* PrMEAN in the same range: PrACRP showed the best linearity with a R^2 of 0.998.

A)



B)

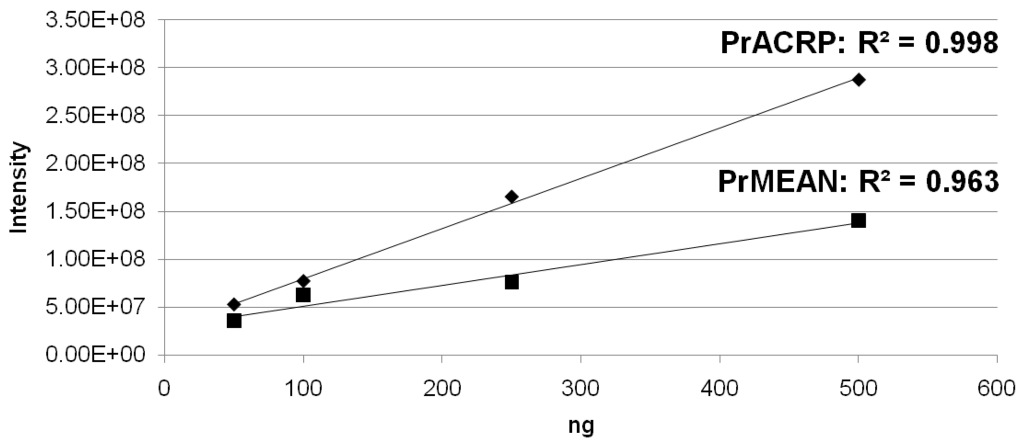


Figure 2.2: Linear correlation of γ-conglutin quantitative parameters *vs.* amount of protein extract tryptic digest (ng) loaded on HPLC-Chip: A) Correlations of the peptide intensities of CRPs (Common Reproducible Peptides) (i.e. 82, 77 and 73); B) Correlations of PrMEAN (Protein Mean Peptide Spectral Intensity) and PrACRP (Protein Average of Common Reproducible Peptide).

It is therefore possible to affirm that the LOQ of γ-conglutin according to the parameter PrACRP is equal to 50 ng of protein extract. Table 2.4 reports the peptide intensity averages, the SDs, and the corresponding RSD % of γ-conglutin CRPs after triplicate analyses of the four protein extracts. In spite of the peptide RSD % might be greater than 20 % in some cultivars, the corresponding RSD % of PrACRPs ranged between 10 and 20 %, which may be considered a good result for a bioanalytical measurement. On the contrary, the RSD % of PrMEANs ranged between 4 and 37 % (Table 2.4).

Table 2.4: The CRPs (Common Reproducible Peptides) of the γ-conglutin in the four cultivars (*cv.* Adam, Arés, Lucky, Multitalia) are reported. For each peptide the index number, the sequence, the average intensity in the three replicates and the corresponding standard deviation and relative standard deviation percentage (RSD %) are reported. The averages of the PrACRP (Protein Average of Common Reproducible Peptides) and of the PrMEAN (Protein Mean Peptide Spectral Intensity), the standard deviations and relative standard deviation percentages (RSD %) are reported too.

Index	CRP γ-conglutin	Adam			Arés			Lucky			Multitalia		
		PeSI average	±SD	RSD%	PeSI average	±SD	RSD%	PeSI average	±SD	RSD%	PeSI average	±SD	RSD%
82	**AGIALGTHQLEENLVVFDLAR**	5.00E+07	7.91E+06	15.83	5.52E+07	3.70E+06	6.70	3.53E+07	2.05E+07	58.04	3.45E+07	1.00E+07	29.01
77	**HSIFEVFTQVFANNVPK**	5.83E+07	1.93E+07	33.13	4.38E+07	7.89E+06	18.03	1.02E+08	1.72E+07	16.89	6.32E+07	8.46E+06	13.38
73	**TPLMQVPVLLDLNGK**	1.80E+07	8.75E+06	48.55	4.07E+07	1.94E+07	47.54	2.60E+07	2.63E+06	10.11	2.15E+07	6.82E+06	31.74
	PrACRP	4.21E+07	8.32E+06	19.77	4.65E+07	4.72E+06	10.13	5.44E+07	1.16E+07	21.39	4.47E+07	7.54E+06	16.86
	PrMEAN	3.64E+07	5.11E+06	14.04	3.41E+07	2.36E+06	6.93	4.41E+07	1.84E+06	4.19	2.70E+07	9.96E+06	36.84

The PrMEANs and PrACRPs of the internal standard BSA (Table 2.3) were used to normalize the corresponding parameters of γ-conglutin: the averages of normalized parameters are shown in Table 2.5.

Table 2.5: Differential analysis of γ-conglutin. The comparison of the N-ACRPs (Normalized Protein Average of Common Reproducible Peptides) and N-MEANs (Normalized Protein Mean Peptide Spectral Intensity) in the four cultivars (*cv.* Adam, Arés, Lucky, Multitalia) is reported. For each cultivar the replicate values (X1, X2, X3), the corresponding averages, standard deviations and relative standard deviation percentages (RSD %) are reported.

		X_1	X_2	X_3	average	± SD	RSD%
γ-conglutin N-ACRP	**Adam**	0.52	0.61	0.69	**0.61**	**0.09**	**14.08**
	Arés	1.06	0.97	0.95	**0.99**	**0.06**	**5.79**
	Lucky	0.75	0.95	1.05	**0.91**	**0.15**	**16.82**
	Multitalia	0.77	0.75	0.82	**0.78**	**0.04**	**4.97**
γ-conglutin N-MEAN	**Adam**	0.64	0.65	0.48	**0.59**	**0.10**	**16.19**
	Arés	0.81	0.63	0.66	**0.70**	**0.10**	**13.67**
	Lucky	0.83	0.60	0.72	**0.72**	**0.12**	**16.15**
	Multitalia	0.43	0.79	0.42	**0.55**	**0.21**	**38.81**

Each N-ACRP parameter had a smaller RSD % than the parent parameter and, moreover, the RSD % of N-ACRP was better than that of N-MEAN, being below 20 %. Some literature data show that the variability of peptide peak areas in different LC-MS analyses can be minimized by integrating small retention time windows of peptide EICs. However, complex mixtures of peptides can produce chromatographic shifts across different samples; in addition, the extension of the chromatographic window increases variability in the quantification of individual peptides for the contribution of coeluting peptides to the peak area.

This analytical problem can be partially solved through a chromatographic alignment. In our investigation, however, the HPLC-Chip showed a high reproducibility of the peptide retention times without any preliminary chromatographic alignment. The RSD % of the retention times of γ-conglutin CRPs ranged between 0.45 and 0.5 %, whereas the RSD % of the retention times of BSA CRPs between 0.98 and 1.69 %. Moreover, it is important to underline that the peptides of γ-conglutin and BSA were well distributed along the chromatogram and covered a wide range of retention times; this assured that the N-ACRP suffered a similar matrix effect in all samples.

2.4.2.3 Proof of the concepts

To evaluate the reliability of the N-ACRP, 100 ng of the protein extract (*cv.* Arés) were spiked with either 10 or 15 ng of BSA to give sample A and B, respectively. Since the amount of γ-conglutin in samples A and B was constant, this protein may be considered as an endogenous internal standard [Wang et al., 2003; Bondarenko et al., 2002; Duranti et al., 1995; Higgs et al., 2005] for the calculation of BSA N-ACRP in these samples (N-ACRP A, N-ACRP B). The experimental ratio N-ACRP B/N-ACRP A was equal to 1.51, very close to 1.50, which was the correct ratio, whereas the ratio between N-MEAN B/N-MEAN A was equal to 2.01. These data demonstrated that in the case of γ-conglutin, NACRP has a better reliability than N-MEAN. N-ACRP may be, thus, adequate for the differential analysis of γ–conglutin: it is reliable and it provides a level of precision close to those obtained using strategies relying on chemical or metabolic labeling [Cutillas et al., 2004; Olsen et al., 2006; Cutillas et al., 2007]. Since sample handling in proteome measurements is highly complex, proteome quantification requires statistical approaches. A common statistical method for testing the equality of means among multiple samples is one-way analysis of variance (one-way ANOVA). N-ACRPs (Table 2.5) were thus submitted to one-way ANOVA (*p* 0.05): this parameter discriminated cultivars with respect to their amount in γ-conglutin. Moreover, the Sheffè test indicated that the cultivar Adam contains less γ-conglutin than the cultivars Multitalia, Arés, and Lucky. In detail, the ratio between the amounts of γ-conglutin was from 1.3 to 1.6 greater in cultivars Multitalia, Arés and Lucky than in cultivar Adam.

2.4.2.4 Quantitative analysis of the vicilins

There are two main bio-analytical limits that complicate the label-free comparative analysis of lupin vicilins: the very complex nature of this heterogeneous class of storage proteins, both because they have a multigenic origin and undergo extensive post-transcriptional modifications [Freitas et al., 2007]; the availability of the homologous precursor sequences deposited in the database until now (*i.e.*, the γ-conglutin precursor and vicilins like protein) may not be exhaustive. Table 2.6 reports all reproducible vicilin peptides identified in the analyses of the four cultivars and classified as CRPs or SRPs: for each peptide the averages of peptide intensities, the SDs and the corresponding RSD % are reported. Table 2.6 reports the averages, the SDs and the RSD % of both N-MEANs and N-ACRPs too.

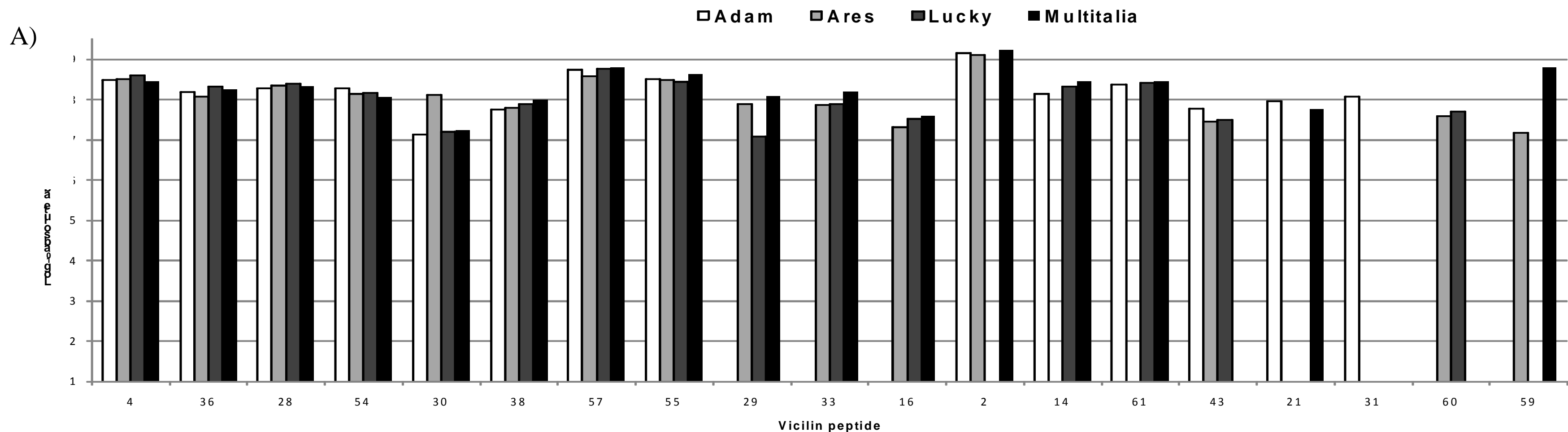
Table 2.6: The CRPs (Common Reproducible Peptides, in grey) and the SRPs (Specific Reproducible Peptides) of the vicilins in the four cultivars (*cv.* Adam, Arés, Lucky, Multitalia) are reported. For each peptide the index number, the sequence, the average intensity in the three replicates and the corresponding standard deviation and relative standard deviation percentage (RSD %) are reported. The comparison of the N-ACRPs (Normalized Protein Average of Common Reproducible Peptides) and N-MEANs (Normalized Protein Mean Peptide Spectral Intensities) in the four cultivars (*cv.* Adam, Arés, Lucky, Multitalia) is reported.

Index		Vicilin	Adam			Arés			Lucky			Multitalia		
			average	± SD	RSD%	average	± SD	RSD%	average	± SD	RSD%	average	± SD	RSD%
4	**CRP**	**HSDADYVLVVLNGR**	**3.12E+08**	**2.92E+07**	**9.36**	**3.20E+08**	**4.70E+07**	**14.71**	**3.92E+08**	**2.50E+07**	**6.38**	**2.86E+08**	**7.22E+07**	**25.24**
36	**CRP**	**INEGALLLPHYNSK**	**1.51E+08**	**4.30E+07**	**28.48**	**1.21E+08**	**2.17E+07**	**17.93**	**2.09E+08**	**1.10E+07**	**5.26**	**1.81E+08**	**2.01E+07**	**11.11**
28	**CRP**	**LAIPINNPGYFYDFYPSSTK**	**1.96E+08**	**4.07E+07**	**20.81**	**2.27E+08**	**1.60E+07**	**7.08**	**2.53E+08**	**4.33E+07**	**17.13**	**2.13E+08**	**1.71E+07**	**8.01**
54	**CRP**	**LLGFGINADENQR**	**1.90E+08**	**5.15E+07**	**27.17**	**1.42E+08**	**3.04E+07**	**21.50**	**1.44E+08**	**3.12E+07**	**21.70**	**1.17E+08**	**3.46E+07**	**29.57**
30	**CRP**	**NTLEATFNTR**	**1.34E+07**	**3.21E+05**	**2.40**	**1.30E+08**	**3.51E+07**	**26.96**	**1.56E+07**	**2.43E+06**	**15.56**	**1.76E+07**	**3.44E+06**	**19.54**
38	**CRP**	**AIFIVVVGEGNGK**	**5.69E+07**	**4.20E+06**	**7.38**	**6.42E+07**	**1.61E+07**	**25.00**	**7.65E+07**	**8.81E+06**	**11.52**	**9.96E+07**	**8.16E+06**	**8.19**
57	**CRP**	**NFLAGSEDNVIR**	**5.42E+08**	**5.95E+07**	**10.98**	**3.78E+08**	**5.06E+07**	**13.38**	**5.93E+08**	**4.68E+07**	**7.89**	**6.20E+08**	**1.16E+08**	**18.65**
55	**CRP**	**LLGFGINAYENQR**	**3.23E+08**	**6.69E+07**	**20.73**	**3.03E+08**	**3.67E+07**	**12.09**	**2.78E+08**	**7.79E+07**	**28.04**	**4.38E+08**	**9.61E+07**	**21.95**
21	**SRP**	IPAGSTSYILNPDDNQKLR	9.29E+07	1.32E+07	14.20	-	-	-	-	-	-	5.80E+07	5.30E+06	9.14
31	**SRP**	NTLEATFNTRYEEIQR	1.22E+08	6.03E+06	4.95	-	-	-	-	-	-	-	-	-

14	**SRP**	QAYNLEYGDALR	1.41E+08	2.41E+07	17.07	-	-	-	2.17E+08	7.66E+07	35.30	2.78E+08	4.92E+07	17.73
2	**SRP**	IVEFQSKPNTLILPK	1.42E+09	8.66E+07	6.10	1.28E+09	7.07E+08	55.19	-	-	-	1.73E+09	9.29E+07	5.36
29	**SRP**	DQQSYFSGFSR	-	-	-	7.64E+07	1.99E+07	26.05	1.21E+07	5.03E+05	4.17	1.22E+08	2.86E+07	23.44
33	**SRP**	ILLGNEDEQEYEEQR	-	-	-	7.37E+07	1.83E+07	24.87	7.69E+07	4.15E+07	53.92	1.60E+08	7.30E+07	45.52
60	**SRP**	AVNELTFPGSAEDIER	-	-	-	4.01E+07	1.32E+07	32.83	5.21E+07	5.56E+06	10.67	-	-	-
59	**SRP**	NFLAGSKDNVIR	-	-	-	1.55E+07	4.10E+06	26.52	-	-	-	6.20E+08	1.16E+08	18.65
61	**SRP**	ELTFPGSAEDIER	2.40E+08	7.78E+07	32.45	-	-	-	2.60E+08	3.95E+07	15.22	2.84E+08	9.39E+07	33.03
39	**SRP**	AIFIVVVGEGNGKYELVGIR	6.02E+07	1.54E+07	25.65	2.85E+07	1.58E+07	55.38	3.23E+07	6.58E+06	20.37	-	-	-
16	**SRP**	qAYNLEYGDALR	-	-	-	2.02E+07	5.68E+06	28.11	3.40E+07	1.20E+07	35.41	3.95E+07	1.27E+07	32.10
-		**N-MEAN**	2.34	0.16	6.90	2.46	0.34	13.76	2.18	0.54	24.29	3.56	0.27	7.58
-	**-**	**N-ACRP**	3.23	0.46	14.16	4.51	0.42	9.32	4.19	0.57	13.50	4.32	0.28	6.52

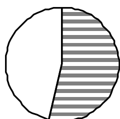
The RSD % of N-MEANs ranged between 7 and 25 %, whereas the RSD % of N-ACRPs ranged between 6 and 14 %. The variability of the NACRP is lower than the variability of N-MEAN because PrACRP is less dependent on peptide responses in ESI than PrMEAN. However, the NACRP is not the most suitable parameter for a correct differential analysis of vicilins. In fact, the vicilin SRPs demonstrated the existence of cultivar-specific isoforms that are expressed in a differential way. Figure 2.3A shows the absolute intensity averages of each reproducible peptide in the four cultivars.

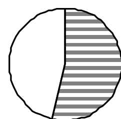
A)

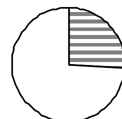


Adam: SRPs 54% **Arés:** SRPs 48% **Lucky:** SRPs 26% **Multitalia:** SRPs 63%

B)







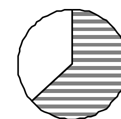


Figure 2.3: A) Log scale plot of the absolute abundance averages of vicilin CRPs (4-55, Common Reproducible Peptides) and SRPs (29-59, Specific Reproducible Peptides) to the total intensity. **B)** Pie-charts of the percentage relative contributions of SRPs (Specific Reproducible Peptides, grey stripes) with respect to the vicilin total intensities calculated as the sum of the absolute abundances of all reproducible peptides in each cultivar.

It was possible to calculate the total intensity in each cultivar as the sum of all peptide absolute intensities. CRPs provide a contribution to the total intensity comparable in each cultivar. Among SRPs, peptide 2 influences drastically the total intensity in the cultivars Adam, Arés and Multitalia providing a percentage contribution ranging from 30 to 40 %. In Figure 3B the analytical weights of all CRPs and all SRPs with respect to the total intensity are reported as percentage ratios. It appears clearly that the analytical weight of SRPs was higher when peptide 2 is included: the percentages of SRPs in cultivars Adam, Arés, and Multitalia were 54, 48 and 63 %, respectively. On the contrary, in cultivar Lucky the percentage of SRPs was only 26 %. The analytical weight of SRPs in cultivar Multitalia (63 %) is the highest due to both the major number of SRPs with respect to the other cultivars and to the high abundance of some of them (peptide 2 and peptide 59). The cultivar Adam has the minor number of SRPs but their analytical weight is important (54 %) due to the presence of peptide 31 that was selectively identified only in this cultivar. Therefore, the quantitative parameter N-ACRP does not appear to be representative of the whole vicilin class because it does not include contribute of the cultivarspecific isoforms. Since we still do not know which peptides are the actual bioactive components, at present the hypocholesterolemic activity may be associated only to the whole vicilin class. For this reason, N-MEAN appears to be the most suitable quantitative parameter for a differential comparison of the whole vicilin class in lupin cultivars.

2.4.2.5 Proof of the concepts

In order to assess the reliability of N-MEAN proposed for the differential analysis of vicilin, 100 ng of the protein extract of the *cv.* Arés was spiked with either 10 ng (sample A) or 15 ng (sample B) of BSA and analyzed after tryptic digestion. The ratio between N-MEANB/N-MEANA was calculated considering the vicilins as the endogenous internal standard as it had already been done for γ-conglutin. The ratio was equal to 1.70, close to the experimental value of 1.50. It seems reasonable to hypothesize that the vicilin N-MEAN is more reliable than the γ-conglutin N-MEAN (1.7 vs. 2.01) because γ-conglutin is a minor protein in the protein extract [Wait et al., 2005] whose concentration is comparable with that of BSA in these samples. On the contrary, vicilins, being among the major seed storage proteins, suffer a minor matrix effect due to the addition of BSA in sample A and B. The N-MEAN results were subjected to the same statistical evaluation as already described for γ-conglutin. One way ANOVA indicated that N-MEAN discriminated the cultivars for their vicilins contents and the Sheffè test indicated that the cultivar Multitalia contains more vicilins than the other cultivars. In detail, the amount of vicilin in cultivar Multitalia was from 1.4 to 1.6 higher than in the other cultivars (Table 2.7).

2.5 Conclusion of study 1

The proposed shotgun-proteomics analysis based on HPLC-Chip-MS/MS allowed a complete characterization of lupin seed storage proteins, since minor proteins, such as γ-conglutin and δ-conglutin, were easily identified together with major proteins with satisfactory coverages without any previous fractionation of the protein extract according to shotgun proteomics workflow. The characterization of each seed storage protein class appears to be as satisfactory as that obtained from the 2-D map analysis in previous works [Wait et al., 2005; Magni et al., 2007].

As concerns the qualitative characterization of lupin proteins, the homology mode search has permitted to identify single aminoacid substitutions in vicilins and δ -conglutin sequences. In particular, the MS/MS sequencing of single aminoacid substituted peptides, starting from the vicilin-like protein and the β-conglutin precursor, enabled to acquire important knowledge of the heterogeneous nature of vicilin isoforms.

In the study 1, the internal standard label-free method based on HPLC-Chip-MS/MS is able to profile the different expression of vicilins and γ-conglutin, potentially, in an unlimited number of protein extracts. Two algorithms were proposed as suitable for the differential analysis of the whole vicilin protein class and the mature γ-conglutin. N-MEAN was the most suitable parameter for profiling the differential expression of the vicilin class, since, being the mean peak intensity of all peptide precursor ions identified; it permits to take into consideration all vicilin isoforms identified in each sample. This is particularly important because there is not yet any clear indication about which peptides are responsible for the hypocholesterolemic activity. Although the difference in the expression of the vicilins in these four cultivars does not appear to be very large, this is the first study showing that lupin cultivars are not perfectly interchangeable for the contents of this bioactive protein class.

Moreover, lupin proteins may be used in the formulation of several functional foods. The functional foods have to contain the proper balance of bioactive ingredients in order to assure their positive impact on the consumer's health in addition to their nutritive value. The HPLC Chip- MS/MS method, based on the addition of the internal standard in complex protein extracts, may be used in the future for the qualitative and quantitative screening of the bioactive vicilin class in functional foods.

The normalized parameter N-ACRP, instead, appears to be the most suitable for profiling the differential expression of γ-conglutin. It was shown to have a good reliability and a precision close to those obtained using strategies relying on chemical or metabolic labeling. It is important to underline that in this work a preliminary chromatogram alignment was not necessary for the high reproducibility of the peptide retention times assured by the HPLC-Chip system. γ-Conglutin has been indicated as a major lupin allergen. The quantification of food allergens is generally based on immunoassays: these methodologies, however, have some limitations, such as the possibility of cross-reactivity with other food proteins or false-positive results. Considering that γ-conglutin has been demonstrated to give cross-reactivity with peanuts allergens [Magni et al., 2004; Breiteneder et al., 2005], it may be very useful to develop a quantitative method that does not rely on the use of antibodies.

The observed linearity and sensitivity suggest that the CRPs may be the base for the development of a Multiple Reaction Monitoring (MRM) method [Anderson, 2006] that may increase significantly the sensitivity of the present label-free method in the detection and quantification of this allergen in complex foods. Such method has been developed in the study 2 thanks to the preliminary differential observations about the γ-conglutin in this study.

2.6 References

Anderson L, Hunter CL, 2006, Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol Cell Proteomics 5:573-588.

Bondarenko PV et al., 2002, Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. Anal Chem 74:4741-4749.

Bradford M, 1976, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248.

Breiteneder H, Mills ENC, 2005, Molecular properties of food allergy. J Allergy Clin Immunol 14-22.

Cutillas PR et al., 2004, The urinary proteome in Fanconi syndrome implies specificity in the reabsorption of proteins by renal proximal tubule cells. Am J Physiol 287:F353-F364.

Cutillas PR, Vanhaesebroeck B, 2007, Quantitative profile of five murine core proteomes using label-free functional proteomics. Mol Cell Proteomics, 1560-1573.

Duranti M et al., 1981, The seed globulins of *Lupinus albus*. Phytochemistry 20:2071-2075.

Duranti M et al., 1995, Lectin-like activity of lupin seed conglutin γ, a glycoprotein previously referred to as a storage protein. J Exp Bot 46:725-728.

Freitas LR et al., 2007, Vicilin-type globulins follow distinct patterns of degradation in different species of germinating legume seeds. Food Chem 102:323-329.

Gallardo K et al., 2001, Proteomic analysis of Arabidopsis seed germination and priming. Plant Physiol 126:835-848.

Gallardo K et al., 2002, Proteomics of Arabidopsis seed germination. A comparative study of wild-type and gibberellin-deficient seeds. Plant Physiol 129:823-837.

Gallardo K et al., 2003, Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiol 133:664-682.

Hajduch M et al., 2005, A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimentional reference maps, expression profiles, and interactive proteome database. Plant Physiol 137:1397-1419.

Higgs RE et al., 2005, Comprehensive label-free method for the relative quantification of proteins from biological samples. J Proteome Res 4:1442-1450.

Locati D et al., 2006, Preliminary approaches for the development of a high performance liquid chromatography/electrospray ionization tandem mass spectrometry method for the detection and label/free semiquantitation of the main storage proteins of *Lupinus albus* in foods. Rapid Commun Mass Spectrom 20:1305-1316.

Magni C et al., 2004, Conglutin gamma, a lupin seed protein, binds insulin in vitro and reduces plasma glucose levels of hyperglycemic rats. J Nutr Biochem 15:646-650.

Magni C et al., 2007, Combined 2-D electrophoretic approaches for the study of white lupin mature seed storage proteome. Phytochemistry 68:997-1007.

Majoul T et al., 2003, Proteomic analysis of the effect of heat stress on hexaploid wheat grain: Characterization of heat-responsive proteins from non prolamins fraction. Proteomics 3:175-183.

Olsen JV et al., 2006, Global, in vivo, and site-specific phosphorylation dynamics in signalling networks. Cell 127:635-648.

Tabata T et al., 2007, Pseudo internal standard approach for label-free quantitative proteomics. Anal., Chem 79:8440-8445.

Wait R et al., 2005, Analysis of *Lupinus albus* storage proteins by two-dimensional electrophoresis and mass spectrometry. J Agric Food Chem 53:4599-4606.

Wang W et al., 2003, Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal Chem 75:4818-4826.

3. Study 2:

Label-free absolute quantification of the γ-conglutin in lupin flour: development and optimization of a nano-HPLC-Chip Multiple Reaction Monitoring (MRM) method based on proteotypic peptides.

3.1 Aim of study 2

In order to estimate the potential health benefit of any dietary supplement or functional food, it is indispensable to have the possibility to get the absolute quantification of each bioactive component.

This study presents an innovative HPLC-Chip-Multiple Reaction Monitoring (MRM) labelfree method for the absolute quantification of γ-conglutin in a total protein extract from lupin flour, i.e. the matrix. This protein is a main bioactive lupin protein, which is very peculiar since, at the same time, it is a hypoglycemic agent and a major allergen in this seed. This research is justified by the increasing interest of food industry for this grain legume, characterized by a high nutritional value and good technological flexibility as well as by the presence of several bioactive proteins, γ-conglutin included. The four main features of the method are the following:

a) the chromatographic separation was performed on a very efficient HPLC-Chip system coupled with a tridimensional Ion Trap mass spectrometer;

b) five proteotypic peptides of γ-conglutin were selected and detected in a Multiple Reaction Monitoring (MRM) mode;

c) the absolute quantification was obtained by the “standard addition” of a highly purified sample of γ-conglutin;

d) the matrix effect was overcome by the addition of a known amount of an exogenous protein, i.e. Bovine Serum Albumin (BSA).

A very critical issue, when developing new HPLC-ESI-MS methods, is the “matrix effect”, i.e. the fact that the detection of each analyte is impaired by other co-eluting components of the matrix. The originality of the present methodology consists in the translation of the “standard addition” strategy from the analytical area to proteomics. The “standard addition” strategy is a common approach used to achieve the absolute quantification of small organic molecules, such as shellfish poisonous toxins [Ito et al., 2001] or essential oils component [Maggi et al., 2009], in the presence of a complex matrix. By using this straightforward approach, in study 2 a reliable absolute quantification of γ-conglutin was reached since the matrix effect was first of all easily estimated thanks to the standard addition approach and consequently minimized using an exogenous internal standard, i.e. Bovine Serum Albumin (BSA) as in study 1.

The standard addition approach is useful both to reach an absolute quantification of the analyte both to evaluate and visualize the matrix effect. In practice, it consists in preparing two different calibration curves of the analyte: the standard curve and the in-matrix curve. The standard curve is obtained by analyzing increasing concentrations of the standard analyte in aqueous solutions by LC-MS. The in-matrix curve is obtained by spiking increasing amounts of the standard analyte in the matrix. The matrix is the heterogeneous sample in which the analyte has to be quantified. The absolute amount of the analyte in the matrix may be calculated from the intercept of the in-matrix curve with the X-axis. Moreover, by comparing slopes of the standard curve and of the in-matrix curve, it is possible to visually evaluate the effect exerted by the matrix on the LC-MS detection of the analyte. More slopes appear to be different and more the matrix effect is important. This means that the coeluting matrix components greatly affected the detection of the analyte. Ideally, two parallel curves suggest the absence of matrix effect.
Figure 3.1 reported the flow scheme for the development and the optimization of the label-free absolute quantification of the γ-conglutin**.**

3.2 Flow scheme of study 2

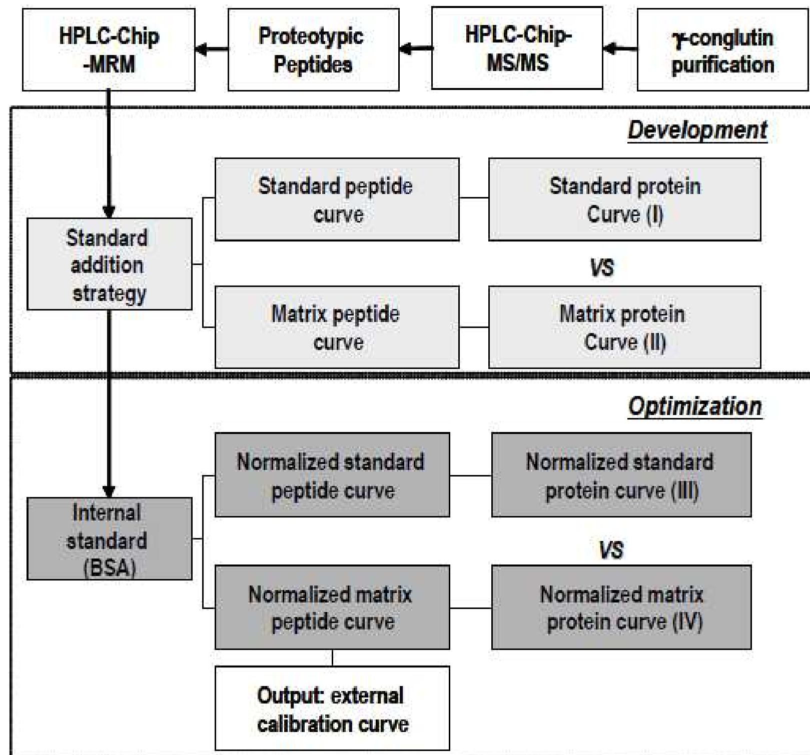


Figure 3.1: Flow scheme for absolute label-free quantification of γ-conglutin: γ-conglutin was purified from raw lupin seed flour; its tryptic digest was analyzed at decreasing concentration in order to individuate the best proteotypic peptides; an HPLC-Chip-Multiple Reaction Monitoring (MRM) method was develop for the selective detection of the γ-conglutin proteotypic peptides. During the development step of the absolute quantitative method the standard addition strategy was used to obtain the standard protein curve (I) and the matrix protein curve (II); in the optimization step an exogenous internal standard protein (BSA) was employed with the aim to solve the matrix effect which affected the method by preparing the normalized standard protein curve (III) and the normalized matrix protein curve (IV).

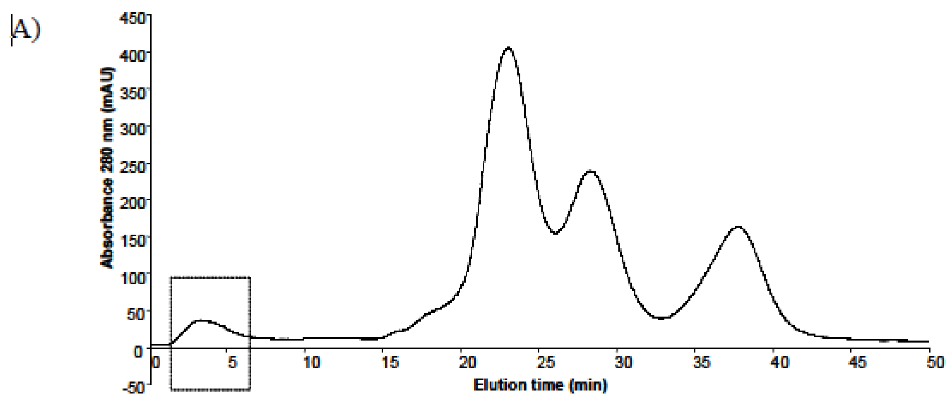
The γ-conglutin was purified from a total lupin seed protein extract and five proteotypic peptides were chosen and used in the HPLC-Chip-MRM method to build the standard protein curve (I) and the in-matrix protein (II) curve in the development phase. In order to minimize the matrix effect that affected the method in the optimization phase a normalized standard protein curve (III) and a normalized in-matrix protein curve (IV) were prepared. The final output of the study is a robust and reliable external calibration curve, i.e. the normalized standard protein curve (III) for the routine quantification of the bioactive γ-conglutin in raw material such as flour.

3.3 Materials and Methods

3.3.1 γ-conglutin purification

Since a standard sample of γ-conglutin is not commercially available, this protein was purified by two chromatographic steps, i.e. anion exchange chromatography followed by gel-filtration chromatography, starting from the lupin total protein extract, obtained by defatted lupin kernel meal with the same procedure described in the study 1. The desalted total protein extract was loaded onto a DEAE-FF column (1.6 x 2.5 cm, 15-70 mm bead size, 5 ml column volume; GE Healthcare Bio-Sciences AB, Sweden). Fractions were eluted from the column with a linear salt gradient (0-100 % NaCl over 16 column volumes) collecting every 30 sec. The γ-conglutin enriched fraction eluted at the beginning of the gradient because the protein was not retained by the stationary phase contrary to all other lupin proteins, which eluted at greater salt concentration. The γ-conglutin enriched fraction was collected and subsequently loaded onto a gel-filtration column (10 x 300 mm, 24 mL column volume; GE Healthcare Bio-Sciences AB, Sweden) for the second purification based on the molecular weight. The calibration of the column was performed using a kit (GE Healthcare Bio-Sciences AB, Sweden) containing the following standard protein: ovalbumin, conalbumin, aldolase, ferritin, thyroglobulin and bluedextran. The molecular weight of the purified γ-conglutin, calculated using the obtained calibration curve ($y = -0.1394 Ln(x) + 1.9809$; $R^2 = 0.9872$) well agrees with the expected molecular weight of native γ-conglutin [Magni et al., 2005]. The purity of the γ-conglutin gel-filtration fraction was established by 2-dimensional (2D) gel electrophoresis under denaturant conditions. For the 2D gel electrophoresis, 40 μg of the purified protein were diluted in IEF solubilization buffer, reduced with 10 mM DTT and alkylated with 20 mM IAM prior to the isoelectric focusing on 7 cm, pH 3-10 non linear IPG stips (Biorad). The second dimension separation was performed on 13 % SDS page gel using Mini Protean 3 Dodeca-cell (Biorad). The gel was stained using Bio-safe Coomassie. As it is possible to see in the Figure 3.2, only the spots of the γ-conglutin large subunit (spot 2) and small subunit (spot 1) were visible.

A)



B)

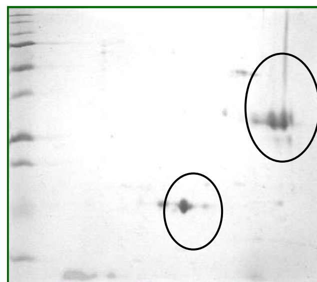


Figure 3.2: Purification of γ-conglutin from the lupin kernel protein extract. Estimation of γ-conglutin purity by twodimensional gel electrophoresis (2DE): only the spots of the γ-conglutin large subunit (spot 2) and small subunit (spot 1) are visible.

3.3.2 HPLC-Chip-MS/MS analysis of the purified γ-conglutin

The purified γ-conglutin was dialyzed against Tris-HCl 100 mM, digested with trypsin at increasing concentration and analyzed via HPLC-Chip-MSMS, acquiring MS/MS spectra in a data-dependent mode, with the aim both to assess the purity of the fraction both to identify the γ-conglutin proteotypic peptides in an experimental manner. Five tryptic digests were prepared as follows: increasing volumes of the dialyzed solution of the purified γ-conglutin were denatured (urea 6 M), reduced by adding 200 mM DTT, alkylated by adding 200 mM IAM and digested with sequencing grade trypsin (0.5 mg/ml) in the ratio 1:50 enzyme/protein (w/w) obtaining tryptic digests with final concentrations reported in Table 3.1.

Table 3.1: Final concentrations of the purified γ-conglutin tryptic digests for HPLC-Chip-MS/MS.

Tryptic digest concentration (ng/µl)	Injection volume (µl)	γ conglutin inject amount (ng)
2.5	2	5
5.0	2	10
10.0	2	20
15.0	2	30
25.0	2	50

The same volume of each digests was injected loading on Chip 5, 10, 20, 30 and 50 ng of γ-conglutin. All experiments were performed on an Agilent 1200 nano-HPLC coupled to an Ion Trap SL series (Agilent Technologies, Palo Alto, CA, US). The ionization source was the HPLC–Chip Cube working in nanoflow electrospray positive ion mode. The chromatographic separation of the tryptic digests was performed on a biocompatible LC Chip containing a 40- nl enrichment column, an analytical column Zorbax 80SB – C18, 5 µm, 0.075 x 43 mm, a capillary tubing connections, a nano-electrospray needle. The HPLC-Chip chromatographic eluents were solvent A: 95 % water, 5 % acetonitrile containing 0.1 % formic acid; solvent B: 95 % acetonitrile, 5 % water containing 0.1 % formic acid; the gradient was: 0 min 3 % B, 50 min 50 % B, 60 min 80 % B, 70 min 3 % B, post run time 5 min at 3 % B; the enrichment of the sample prior to gradient start was performed at 4 µl/min using the loading pump and solvent C: 99 % water, 1 % acetonitrile containing 0.1 % formic acid. The capillary voltage was 1850 V, endplate offset – 500 V, drying gas flow 3 L/min, drying gas temperature 300 °C; the ICC target was 30.000, the maximum accumulation time was 150 ms and 2 spectra were averaged. The data dependent Auto MS(n) parameters were: scan range 300 – 2000 m/z, target mass 700 m/z, MS/MS stage 2, fragmentation amplitude 1 V, precursor ions 2, preferred doubly charge ions. The raw data of the AutoMSn analyses were processed using Spectrum Mill Proteomics Workbench (Agilent Technologies, Palo Alto, California) using the setting parameters described in the study 1. The HPLC-Chip-MS/MS analysis of the γ- conglutin tryptic digest enabled to identify only "conglutin-γ" (NCBInr accession number 11191819) with a cumulative sequence coverage equal to 33 %. These results confirmed the high degree of purity of the purified γ-conglutin fraction showed by 2DE.

3.4 Results

3.4.1 Proteotypic peptides and HPLC-Chip-MRM

An unambiguous protein characterization cannot be based on a single proteolytic peptide, which it is nothing more than a fragment of its precursor protein and could derive simultaneously from multiple precursors. With the same logic, accurate quantification can rarely be based only on one peptide. Since a single peptide only defines a very short segment of a protein, there are two different risks: a) an overestimation of the target protein, when this single peptide is common to two or more similar proteins or protein classes; b) an underestimation of some relevant modified forms in the presence of closely related isoforms (i.e. post-translational modified forms of the protein) that are not recognized by the single peptide. Consequently, the quantification of a target protein based on more than one peptide is much more reliable than that based on a single peptide. Moreover, although proteotypic peptides can be predicted by computation or extracted from proteomic databases [Sanders et al., 2007, Blonder et al., 2007], they need also to be experimentally validated. Considering all these facts, five different proteotypic peptides of γ-conglutin were chosen to develop the label-free quantitative method using an experimental observational frequencies. In practice, increasing concentrations of the tryptic digest of the purified γ-conglutin (Table 3.1) were loaded on the HPLC-Chip and analyzed in data-dependent full scan mode. Only those peptides that were detectable even at the lowest concentration were chosen as proteotypic peptides confirming their optimal features in term of chromatographic behavior and mass spectrometric detectability. Another important feature for selection was their uniqueness for the γ-conglutin sequence among all lupin proteins reported in the NCBInr database. This was not difficult because the γ-conglutin sequence is the most peculiar one among lupin proteins, having a very low homology either with the vicilins or the legumins. In addition, the previous study has demonstrated that the sequence of this protein is highly conserved in lupin, since different cultivars of *Lupinus albus* do not present any single aminoacid substitution along the γ-conglutin sequence. In addition, their retention time distribution along the gradient was considered while selecting those to be included in the quantitative methods. Table 3.2 reported the sequence, the retention time, the charge (z), the mass to charge ratio (m/z) of the precursor ions of the five proteotypic peptides. The five proteotypic peptides assured a satisfactory sequence coverage of γ-conglutin (15 %). Their retention times were well distributed along the elution gradient, enabling the detection by mass spectrometry in two different segments: the peptides 1 and 2 (Table 3.2) were detected in the first segment (0 – 18 min), and the peptides 3, 4, and 5 in the second (18 – 35 min).

Table 3.2: Proteotypic peptides of γ-conglutin: entry, amino acid sequence, retention time (Rt), charge state (z), mass to charge ratios (m/z) of each precursor ion and m/z ratios of the main product ions used for the quantification of each precursor ion.

Entry	**Sequence**	**Rt (min)**	**Precursor ion (z)**	**Precursor ion (m/z)**	**Product ions (m/z)**
1	VGFNTNSLK	6.3	2	490.2	676.4, 562.3, 481.3
2	SCSNLFDLNNP	14.6	2	640.8	572.2, 457.2, 709.3
3	IPQFLFSCAPTFLTQK	24.1	2	633.4	834.5, 487.3, 417.7
4	AGIALGTHQLEENLVVFDLAR	23.1	3	756.2	774.0, 720.5, 621.4
5	HSIFEVFTQVFANNVPK	28.2	3	659.9	594.9, 789.6, 867.7

The mass spectrometric detection of the proteotypic peptides was developed in MRM mode, since this is more sensitive and specific than the data-dependent full scan mode, in which a subset of high signal peptides seen in the first MS stage (MS1) is subjected to the second MS/MS stage (MS2). In the MRM mode, only specific parent ions are selectively monitored along all the analysis. These ions are isolated and fragmented into the ion trap during alternated cycles of MS1 and MS2, which are sequentially repeated for each parent ion.
Consequently, the MRM approach provides a high structural specificity for the target peptides that are chosen as representatives of the cleaved protein and moreover it maximizes the sensitivity of the ion trap analyzer partially overcoming its intrinsic limitation related to the limited capability. The multiple reaction monitoring (MRM) acquisition of the proteotypic peptide spectra was divided into two temporal segments; in segment 1 (0 - 18 min) the precursor ions 490.24 m/z, 582.32 m/z for γ-conglutin were selectively monitored; in segment 2 (18 - 35 min) the precursor ions 633.4 m/z, 756.2 m/z, and 659.9 m/z for γ-conglutin were monitored. The "QuantAnalysis" data package was used to build calibration curves for each precursor ion after the MRM data acquisition (i.e. peptide curves). The calibration curves were build by integrating at each calibration level the signal of the three main product ions in the fragmentation spectra of each precursor ion. These three product ions were the most intense among all the product ions in each MS/MS spectrum. Table 3.2 reports the selected product ion mass to charge (m/z) ratios for each precursor ion.

3.4.2 Development: standard addition strategy

The "standard addition" method consists in the comparison between the curve of the standard analyte dissolved in a suitable solvent, i.e. the standard curve, and the curve of the same analyte spiked into the samples in which it has to be quantified, i.e. the in-matrix curve. It enables both the absolute quantification of the analyte, calculated from the intercept of the "in-matrix curve" with the X-axis, and an estimation of the matrix effect by comparing the slopes of the "standard curve" and "in-matrix curve". In the presence of a negligible matrix effect, the two curves are parallel, whereas the slopes are different when the matrix has a major effect on the detection of the analyte. The originality of the study 2 consists in the translation of the "standard addition" strategy from the analytical area to label-free quantitative proteomics. According to the standard addition strategy a standard protein curve and an in-matrix protein curve were prepared for the quantification of the γ-conglutin.
Samples for the standard protein curve were obtained by digesting increasing amounts of the purified γ-conglutin in absence of the matrix, i.e. the total protein extract obtained by the defatted lupin flour containing an unknown amount of endogenous γ-conglutin to be quantified (Table 3.3). Briefly, increasing volumes of the dialyzed solution of the purified γ- conglutin were denatured (urea 6 M), reduced by adding 200 mM DTT, alkylated with 200 mM IAM and digested with sequencing grade trypsin (0.5 mg/ml) in the ratio 1:50 enzyme/protein (w/w) obtaining tryptic digests with the same final volume and with the concentrations reported in Table 3.3. Six calibration levels were used for preparing the standard protein curve; 2 μl of each calibration level (Table 3.3) were injected in order to load on the chip 0, 5, 10, 20, 30, and 50 ng of purified γ-conglutin. Analyses were performed in triplicate.

Table 3.3. Description of standard protein curve samples: concentrations of the tryptic digests (ng/µL), injection volumes (I.V.), amounts of purified γ-conglutin and matrix loaded on the HPLC-Chip column at each calibration level (C.L).

Standard protein curve				
C.L.	**Concentration (ng/µl)**	**I.V (µl)**	**Cg (ng)**	**matrix**
1	0.0	2	0	0
2	2.5	2	5	0
3	5.0	2	10	0
4	10.0	2	20	0
5	15.0	2	30	0
6	25.0	2	50	0

In Table 3.4 the peptide area values of each proteotypic peptide (numbered as 1, 2, 3, 4 and 5 according to the Table 3.2) for each replicate analysis were reported together with the correspondent area average, standard deviation, and RSD % at each calibration level. The RSD % of the peptide 1 ranged from 4.95 to 15.89 %; the RSD % of peptide 2 from 2.54 to 14.74 %, peptide 3 from 2.55 to 12.87 %, peptide 4 from 1.52 to 18.58 % and finally peptide 5 from 1.92 to 13.38 %.

Table 3.5 reported the Replicate areas calculated by averaging the areas of all proteotypic peptide (numbered as 1, 2 , 3, 4, and 5 according to Table 3.2) within replicate 1, 2 or 3 at each calibration level (C.L.); the three Replicate areas at each calibration level were averaged to obtain the calibration level area. In Table 3.5 the standard deviation (sd) and the RSD % of the Calibration Level area (C.L. area) were reported too. The standard protein curve was obtained by plotting the six C.L. areas against the nanograms of purified γ-conglutin loaded on chip at each calibration level. The RSD % of the C.L. areas ranged from 0.92 to 4.57 % demonstrating an high reproducibility of this value at each calibration level.

Table 3.4: Values for the preparation of standard peptide curves. Standard peptide curves are obtained by plotting the peptide area averages (average) of each proteotypic peptide (numbered as 1, 2 , 3, 4, and 5 according to Table 3.2) at each calibration level (C.L) vs. the amount of γ-conglutin (Cγ ng) loaded on HPLC-Chip column. Peptide area averages are obtained by averaging peptide areas (area) of each replicate within each calibration level. The standard deviation (sd) and the RSD % of peptide area averages are reported too.

C.L.	Cγ (ng)	peptide 1				peptide 2				peptide3				peptide 4				peptide 5			
		area	average	sd	RSD%	area	average	sd	RSD%	area	average	sd	RSD%	area	average	sd	RSD%	area	average	sd	RSD%
1	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
		1.63E+07				8.13E+07				3.75E+07				7.67E+07				1.17E+08			
2	5	1.71E+07	1.72E+07	8.95E+05	5.21	7.18E+07	7.11E+07	1.05E+07	14.74	4.35E+07	4.15E+07	3.49E+06	8.41	7.81E+07	7.60E+07	2.58E+06	3.40	1.22E+08	1.30E+08	1.73E+07	13.38
		1.81E+07				6.03E+07				4.36E+07				7.31E+07				1.50E+08			
		3.27E+07				1.54E+08				1.34E+08				1.69E+08				3.93E+08			
3	10	3.52E+07	3.73E+07	5.93E+06	15.89	1.53E+08	1.55E+08	3.95E+06	2.54	1.57E+08	1.47E+08	1.15E+07	7.85	1.71E+08	1.72E+08	2.61E+06	1.52	3.55E+08	3.62E+08	2.75E+07	7.60
		4.40E+07				1.60E+08				1.50E+08				1.75E+08				3.39E+08			
		8.29E+07				2.64E+08				3.74E+08				2.68E+08				6.21E+08			
4	20	8.74E+07	8.30E+07	4.34E+06	5.23	2.47E+08	2.56E+08	8.15E+06	3.19	3.01E+08	3.25E+08	4.19E+07	12.87	2.69E+08	2.71E+08	4.76E+06	1.76	6.43E+08	6.07E+08	4.55E+07	7.50
		7.87E+07				2.56E+08				3.01E+08				2.77E+08				5.56E+08			
		1.24E+08				4.10E+08				5.11E+08				3.80E+08				9.72E+08			
5	30	1.24E+08	1.20E+08	5.96E+06	4.95	4.00E+08	4.07E+08	6.82E+06	1.68	5.37E+08	5.15E+08	1.98E+07	3.84	3.80E+08	4.25E+08	7.90E+07	18.58	8.53E+08	9.27E+08	6.43E+07	6.93
		1.14E+08				4.12E+08				4.98E+08				5.16E+08				9.55E+08			
		1.79E+08				5.84E+08				8.56E+08				7.72E+08				1.80E+09			
6	50	2.05E+08	1.90E+08	1.30E+07	6.82	6.38E+08	5.99E+08	3.37E+07	5.63	8.14E+08	8.35E+08	2.13E+07	2.55	7.35E+08	7.53E+08	1.85E+07	2.46	1.75E+09	1.79E+09	3.43E+07	1.92
		1.87E+08				5.75E+08				8.34E+08				7.51E+08				1.82E+09			

Table 3.5: Values for preparing standard protein curve. Standard protein curve is obtained by plotting calibration level areas (C.L. areas) vs. the amount of γ-conglutin (Cγ ng) loaded on HPLC Chip column. C.L. areas are calculated by averaging the three Replicate areas within each calibration level (C.L.). Replicate areas are obtained by averaging the peptide areas of all the five proteotypic peptides (numbered as 1, 2 , 3, 4, and 5 according to Table 3.2) within each replicate (Rep. N°). Standard deviation (sd) and RSD % of C.L. area are reported too.

C.L.	Cγ (ng)	Rep. N°	peptide					Replicate area	C.L. area	sd	RSD%
			1	2	3	4	5				
1	0	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	5	1	1.63E+07	8.13E+07	3.75E+07	7.67E+07	1.17E+08	6.58E+07	6.71E+07	1.63E+06	2.42
		2	1.71E+07	7.18E+07	4.35E+07	7.81E+07	1.22E+08	6.65E+07			
		3	1.81E+07	6.03E+07	4.36E+07	7.31E+07	1.50E+08	6.89E+07			
3	10	1	3.27E+07	1.54E+08	1.34E+08	1.69E+08	3.93E+08	1.77E+08	1.75E+08	1.60E+06	0.92
		2	3.52E+07	1.53E+08	1.57E+08	1.71E+08	3.55E+08	1.74E+08			
		3	4.40E+07	1.60E+08	1.50E+08	1.75E+08	3.39E+08	1.73E+08			
4	20	1	8.29E+07	2.64E+08	3.74E+08	2.68E+08	6.21E+08	3.22E+08	3.08E+08	1.41E+07	4.57
		2	8.74E+07	2.47E+08	3.01E+08	2.69E+08	6.43E+08	3.10E+08			
		3	7.87E+07	2.56E+08	3.01E+08	2.77E+08	5.56E+08	2.94E+08			
5	30	1	1.24E+08	4.10E+08	5.11E+08	3.80E+08	9.72E+08	4.79E+08	4.79E+08	2.03E+07	4.23
		2	1.24E+08	4.00E+08	5.37E+08	3.80E+08	8.53E+08	4.59E+08			
		3	1.14E+08	4.12E+08	4.98E+08	5.16E+08	9.55E+08	4.99E+08			
6	50	1	1.79E+08	5.84E+08	8.56E+08	7.72E+08	1.80E+09	8.38E+08	8.33E+08	4.48E+06	0.54
		2	2.05E+08	6.38E+08	8.14E+08	7.35E+08	1.75E+09	8.29E+08			
		3	1.87E+08	5.75E+08	8.34E+08	7.51E+08	1.82E+09	8.34E+08			

Samples for the in-matrix protein curve were obtained by digesting with the same procedure described above increasing amounts of the purified γ-conglutin in the presence of a constant amount of the matrix, i.e the protein extract containing all seed storage proteins of lupin among which the endogenous γ-conglutin in an unknown amount: increasing volumes of the dialyzed solution of the purified γ–conglutin were spiked in constant volumes of the protein extract in order to obtain tryptic digest with the final concentrations reported in Table 3.6. Six calibration levels were used also for preparing the in-matrix protein curve; 2 μl of each calibration level (Table 3.6) were injected in order to load on the chip 0, 5, 10, 20, 30, and 50 ng of γ-conglutin in presence of 200 ng of matrix. The analyses were performed in triplicate.

Table 3.6: Description of the in-matrix protein curve samples: concentrations of the tryptic digests (ng/μL), injection volumes (I.V.), amounts of purified γ-conglutin (ng) and matrix (ng) loaded on the HPLC-Chip column at each calibration level (C.L).

In-matrix protein curve *				
C.L.	Concentration (ng/μL)	I.V. (μL)	Cγ (ng)	Matrix (ng)
1	100.0	2	0	200
2	102.5	2	5	200
3	105.0	2	10	200
4	110.0	2	20	200
5	115.0	2	30	200
6	125.0	2	50	200

*In addition to the spiked γ-conglutin, these samples contained an unknown amount of endogenous γ-conglutin inside the matrix.

In Table 3.7 the peptide area values of each proteotypic peptide (numbered as 1, 2, 3, 4 and 5 according to the Table 3.2) for each replicate were reported together with the correspondent area average, standard deviation, and RSD % of the area average at each calibration level. The RSD % of the peptide 1 in presence of the matrix ranged from 1.16 to 9.81 %; the RSD % of peptide 2 from 1.83 to 6.63 %, peptide 3 from 3.92 to 21.27 %, peptide 4 from 1.47 to 16.70 % and finally peptide 5 from 4.23 to 27.12 %.
Table 3.8 reported the Replicate areas calculated by averaging the areas of all proteotypic peptide (numbered as 1, 2 , 3, 4, and 5 according to Table 3.2) within replicate 1, 2 or 3 at each calibration level (C.L.); the three Replicate areas were then averaged to obtain the calibration level areas (C.L. area). In Table 3.8 the standard deviation (sd) and the RSD % of the C.L. area were reported too. The in-matrix protein curve was obtained by plotting the six calibration areas against the nanograms of purified γ-conglutin loaded on chip at each calibration level (i. e. 0, 5, 10, 20, 30 and 50 ng) and in presence of a constant amount of matrix loaded on chip (i.e. 200 ng, see Table 3.6). The RSD % of the C.L. area ranged from 5.31 to 13.55 %.

Table 3.7: Values for the preparation of in-matrix peptide curves. In-matrix peptide curves are obtained by plotting the peptide area averages (average) of each proteotypic peptide (numbered as 1, 2, 3, 4, and 5 according to Table 3.2) at each calibration level (C.L) vs. the amount of γ-conglutin (Cγ ng) loaded on HPLC-Chip column in presence of 200 ng of matrix. Peptide area averages are obtained by averaging peptide areas (area) of each replicate within each calibration level. The standard deviation (sd) and the RSD % of peptide area averages are reported too.

C.L.	Cγ (ng)	peptide 1				peptide 2				peptide3				peptide 4				peptide 5			
		area	average	sd	RSD%	area	average	sd	RSD%	area	average	sd	RSD%	area	average	sd	RSD%	area	average	sd	RSD%
		1.33E+07				4.71E+07				4.24E+07				5.40E+07				1.44E+08			
1	0	1.49E+07	1.35E+07	1.33E+06	9.81	4.91E+07	4.70E+07	2.09E+06	4.44	4.78E+07	4.26E+07	5.07E+06	11.89	5.12E+07	5.10E+07	3.10E+06	6.07	1.50E+08	1.31E+08	2.83E+07	21.64
		1.23E+07				4.49E+07				3.77E+07				4.78E+07				9.84E+07			
		1.91E+07				6.74E+07				6.82E+07				7.25E+07				2.10E+08			
2	5	1.98E+07	1.91E+07	6.63E+05	3.46	6.19E+07	6.47E+07	2.76E+06	4.27	7.08E+07	6.82E+07	2.67E+06	3.92	7.43E+07	6.69E+07	1.12E+07	16.70	2.14E+08	2.07E+08	8.75E+06	4.23
		1.85E+07				6.46E+07				6.55E+07				5.41E+07				1.97E+08			
		2.55E+07				7.28E+07				9.86E+07				7.59E+07				1.97E+08			
3	10	2.58E+07	2.52E+07	8.38E+05	3.32	7.57E+07	7.58E+07	3.09E+06	4.07	9.88E+07	1.01E+08	4.72E+06	4.66	8.32E+07	7.87E+07	3.94E+06	5.00	1.29E+08	1.84E+08	5.00E+07	27.12
		2.42E+07				7.89E+07				1.07E+08				7.70E+07				2.26E+08			
		3.94E+07				1.31E+08				1.61E+08				1.24E+08				4.72E+08			
4	20	4.18E+07	4.12E+07	1.64E+06	3.98	1.30E+08	1.36E+08	8.99E+06	6.63	1.63E+08	1.48E+08	2.50E+07	16.88	1.35E+08	1.31E+08	5.97E+06	4.55	3.86E+08	4.00E+08	6.63E+07	16.56
		4.25E+07				1.46E+08				1.19E+08				1.34E+08				3.42E+08			
		5.39E+07				1.71E+08				1.49E+08				1.55E+08				5.10E+08			
5	30	5.29E+07	5.32E+07	6.17E+05	1.16	1.65E+08	1.69E+08	3.09E+06	1.83	2.24E+08	1.97E+08	4.20E+07	21.27	1.78E+08	1.64E+08	1.19E+07	7.21	4.90E+08	4.59E+08	7.14E+07	15.55
		5.28E+07				1.69E+08				2.19E+08				1.60E+08				3.78E+08			
		7.50E+07				2.09E+08				2.72E+08				2.16E+08				5.44E+08			
6	50	7.33E+07	7.57E+07	2.79E+06	3.69	2.24E+08	2.14E+08	8.71E+06	4.07	3.27E+08	2.95E+08	2.88E+07	9.79	2.18E+08	2.15E+08	3.15E+06	1.47	6.46E+08	6.15E+08	6.20E+07	10.08
		7.88E+07				2.08E+08				2.85E+08				2.12E+08				6.56E+08			

Table 3.8: Values for preparing in-matrix protein curve. In-matrix protein curve is obtained by plotting calibration level areas (C.L. areas) vs. the amount of γ-conglutin (Cγ ng) loaded on HPLCChip column in presence of 200 ng of matrix. C.L. areas are calculated by averaging the three Replicate areas within each calibration level (C.L.). Replicate areas are obtained by averaging the peptide areas of all the five proteotypic peptides (numbered as 1, 2 , 3, 4, and 5 according to Table 3.2) within each replicate (Rep. N°). Standard deviation (sd) and RSD % of C.L. area are reported too.

C.L.	Cγ (ng)	Rep. N°	peptide					Replicate area	C.L. area	sd	RSD%
			1	2	3	4	5				
1	0	1	1.33E+07	4.71E+07	4.24E+07	5.40E+07	1.44E+08	6.01E+07	5.70E+07	7.73E+06	13.55
		2	1.49E+07	4.91E+07	4.78E+07	5.12E+07	1.50E+08	6.27E+07			
		3	1.23E+07	4.49E+07	3.77E+07	4.78E+07	9.84E+07	4.82E+07			
2	5	1	1.91E+07	6.74E+07	6.82E+07	7.25E+07	2.10E+08	8.74E+07	8.52E+07	4.53E+06	5.31
		2	1.98E+07	6.19E+07	7.08E+07	7.43E+07	2.14E+08	8.82E+07			
		3	1.85E+07	6.46E+07	6.55E+07	5.41E+07	1.97E+08	8.00E+07			
3	10	1	2.55E+07	7.28E+07	9.86E+07	7.59E+07	1.97E+08	9.40E+07	9.31E+07	1.01E+07	10.87
		2	2.58E+07	7.57E+07	9.88E+07	8.32E+07	1.29E+08	8.25E+07			
		3	2.42E+07	7.89E+07	1.07E+08	7.70E+07	2.26E+08	1.03E+08			
4	20	1	3.94E+07	1.31E+08	1.61E+08	1.24E+08	4.72E+08	1.86E+08	1.71E+08	1.45E+07	8.45
		2	4.18E+07	1.30E+08	1.63E+08	1.35E+08	3.86E+08	1.71E+08			
		3	4.25E+07	1.46E+08	1.19E+08	1.34E+08	3.42E+08	1.57E+08			
5	30	1	5.39E+07	1.71E+08	1.49E+08	1.55E+08	5.10E+08	2.08E+08	2.09E+08	1.32E+07	6.32
		2	5.29E+07	1.65E+08	2.24E+08	1.78E+08	4.90E+08	2.22E+08			
		3	5.28E+07	1.69E+08	2.19E+08	1.60E+08	3.78E+08	1.96E+08			
6	50	1	7.50E+07	2.09E+08	2.72E+08	2.16E+08	5.44E+08	2.63E+08	2.83E+08	1.76E+07	6.24
		2	7.33E+07	2.24E+08	3.27E+08	2.18E+08	6.46E+08	2.98E+08			
		3	7.88E+07	2.08E+08	2.85E+08	2.12E+08	6.56E+08	2.88E+08			

Figure 3.3 reported standard curves (I) and in-matrix curves (II) for each proteotypic peptides (peptide 1 in plot A, peptide 2 in plot B, peptide 3 in plot C, peptide 4 in plot D, peptide 5 in plot E) of γ-conglutin. Figure 3.4 reported the standard protein curve (I in plot F) and the inmatrix protein curve (II in plot F).
The standard peptide curves (AI, BI, CI, DI, and EI of Figure 3.3) were obtained by plotting peptide area average vs. the amount of the purified γ-conglutin (ng) loaded on chip at each calibration level (Table 3.4). They were characterized by a very good linearity: the regression coefficients were respectively R^2 = 0.995 for peptide 1, R^2 = 0.990 for peptide 2, R^2 = 0.997 for peptide 3, R^2 = 0.992 for peptide 4 and R^2 = 0.997 for peptide 5. The standard protein curve (Figure 3.4, curve FI) was obtained by plotting the calibration level areas vs. the amount (ng) of the purified γ-conglutin loaded on chip at each calibration level (Table 3.5). It had a satisfactory regression coefficient (R^2 = 0.998) and its intercept with the Y-axis was very close to 0.
The in-matrix peptide curves (AII, BII, CII, DII, and EII of Figure 3.3) were obtained by plotting the peptide area averages vs. the amount of the purified γ-conglutin (ng) loaded on chip at each calibration level in presence of a constant amount of matrix (Table 3.7). They showed a slightly worse linearity than the corresponding standard curves, especially in case of the peptide 2 (plot B) and 5 (plot E). Consequently, the in-matrix protein curve (Figure 3.4, curve FII), which essentially may be considered as the average of the five in-matrix peptide curves because it is obtained by plotting the calibration level area vs the nanograms of the purified γ-conglutin loaded on chip in presence of a constant amount of the matrix (Table 3.8), showed a regression coefficient (R^2=0.987) lower than the R^2 of the standard protein curve (Figure 3.4, curve FI).

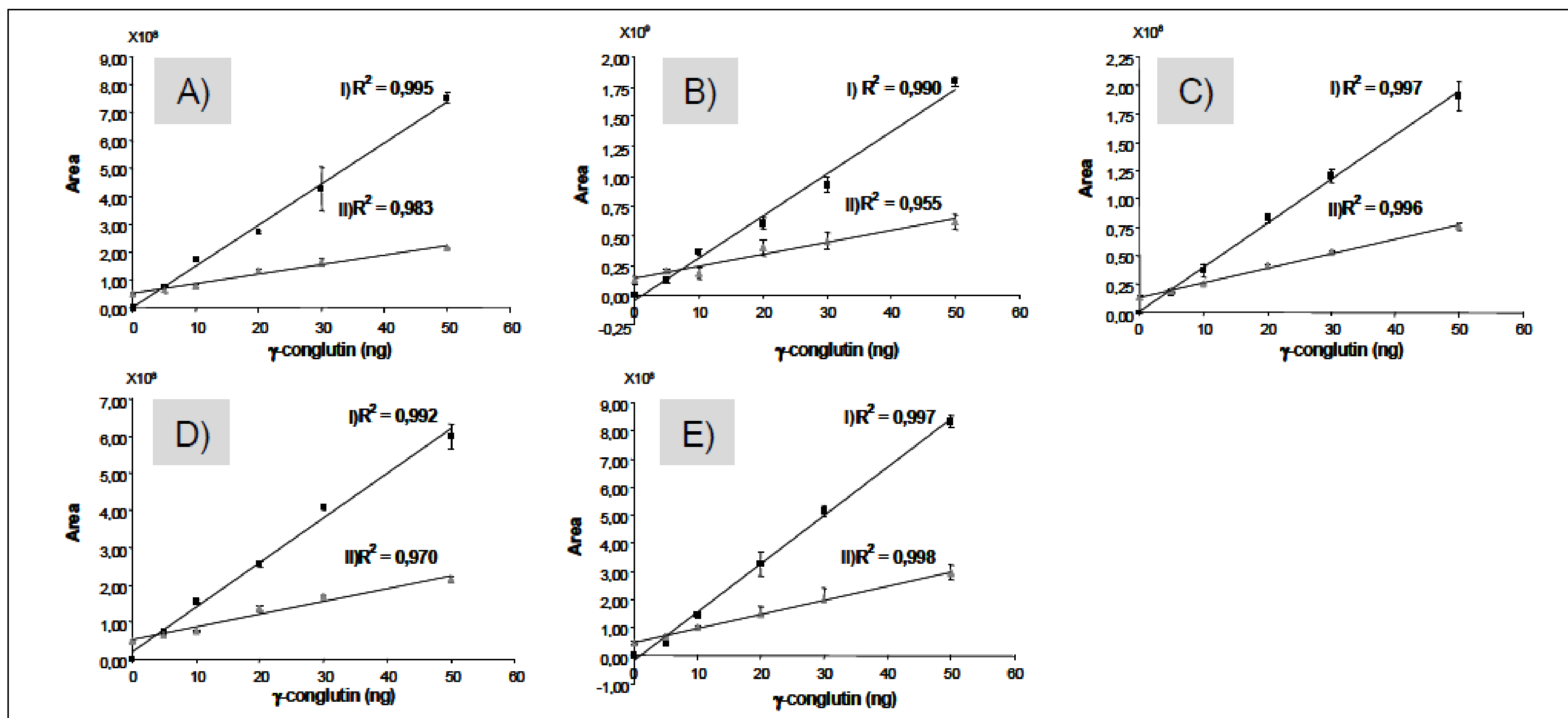


Figure 3.3: Standard peptide curves (I)and in-matrix peptide curves (II) of peptide 1 (plot A), 2 (plot B), 3 (plot C), 4 (plot D) and 5 (plot E).

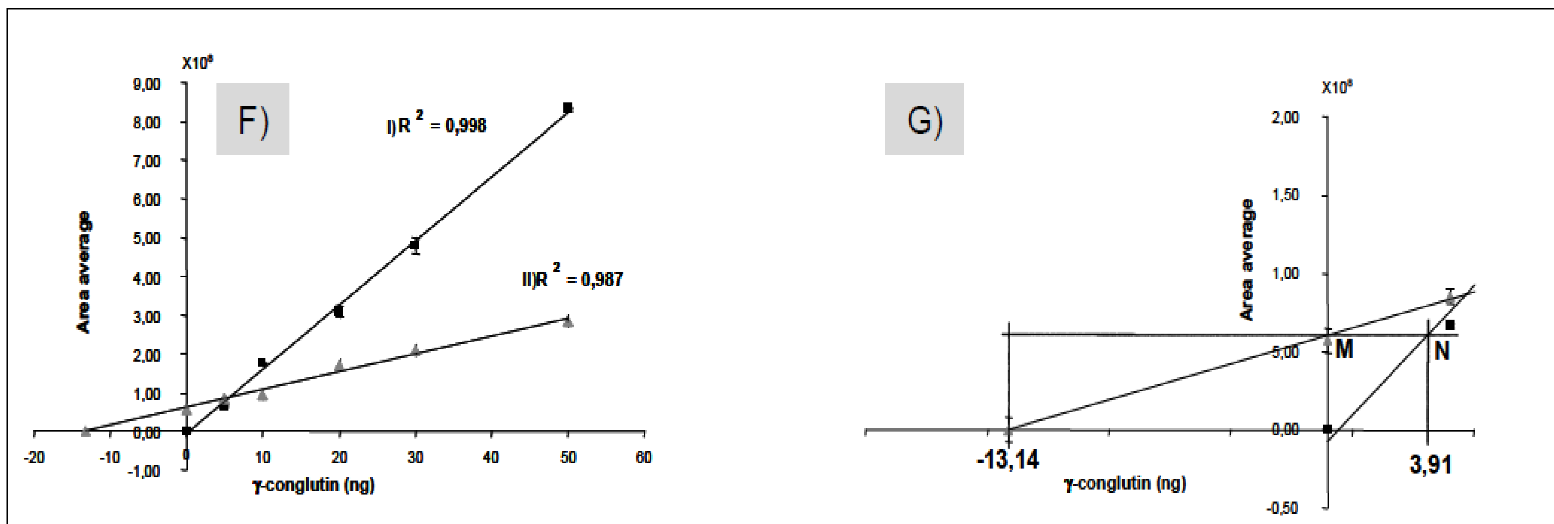


Figure 3.4: Plot F) Standard protein curve (I) and in-matrix protein curve (II); Plot G) Evaluation of the absolute quantification of γ-conglutin obtained from the standard protein curve and in matrix protein curve.

The comparison between the standard and the in-matrix curves enabled to estimate the effect exerted by the matrix on the peptide detection. All the in-matrix peptide curves had smaller slopes than the corresponding standard peptide curves showed a worse linearity (Figure 3.3). This means that the co-eluting matrix components caused an important suppression of the peptide ionization and, consequently, of the MS detectability. This was clearly confirmed by the big differences of the slopes of the in-matrix protein curve and the standard protein curve. In order to make the standard addition approach suitable for a reliable absolute quantitative method in proteomics, it was indispensable to find the way to overcome all these problems.
A number of options have been proposed for minimizing the matrix effect in LC-MS quantitative analysis. The most important are the reduction of the number of the co-eluting compounds by the improvement of the chromatographic separation [Pascoe et al., 2001], the reduction of the flow rate to enhance the sensitivity of the LC-MS platform and the dilution of the sample.
From the beginning this study had been planned for the best, since all the chromatographic separations were performed by using a nano HPLC-Chip system, which enabled both an online enrichment/clean-up step before the chromatographic separation and a drastic reduction of the flow rate (0.3 µL/min) assuring a high sensitivity and a very satisfactory retention time reproducibility (the RSD % of the retention times for both γ-conglutin and BSA peptides ranged between 0.5 % and 1.2 %).
Since the chromatographic improvement had not been sufficient to completely avoid any matrix effect, it was decided to add an exogenous internal standard also with the scope to solve another main problem, which negatively influences the quantification accuracy, the tryptic digestion efficiency. This is a typical problem of the AQUA methods, that do not not take into account the yield of the digestion, since the labeled peptides are added to the samples after the digestion step. In our method, instead, the failure to take into account the actual efficiency of the proteolytic step is compensated by simultaneously co-digesting the added exogenous protein and the purified γ-conglutin, as well as the matrix proteins.

3.4.3 Optimization: exogenous internal standard BSA

Since the chromatographic improvement had not been sufficient to completely avoid any matrix effect, it was decided to use an exogenous internal standard, i.e. BSA. The HPLCChip- MRM method was implemented to detect two BSA peptides chosen to normalize the area of the proteotypic peptides. The BSA peptide LVNELTEFAK (precursor ion 582.3 m/z, charge 2+, 12.5 minutes of retention time, main product ion 494.3, 595.4, 951.5 m/z) was chosen to normalize areas of γ-conglutin peptides 1 and 2 in the segment 1; the peptide DAFLGSFLYEYSR (precursor ion 784.5 m/z, charge 2+, 22.7 minutes of retention time, product ion 717.4, 775.4, 1121.7 m/z) to normalize areas of γ-conglutin peptides 3 ,4 and 5 in the segment 2.
Samples for the normalized standard protein curve were obtained by digesting with the same experimental procedure previously described, increasing amounts of the purified γ-conglutin in the presence of a constant amount of the internal standard BSA: samples were spiked with the same volume of a BSA standard solution prior to the digestion step in order to obtain tryptic digests with the final concentration reported in Table 3.9. All samples were analyzed in three replicates using the implemented HPLC-Chip-MRM method by injecting 2 µl of each sample.

Table 3.9: Description of normalized standard protein curve samples: concentrations of the tryptic digests (ng/µL), injection volumes (I.V.), amounts of purified γ-conglutin and matrix loaded on the HPLC-Chip column at each calibration level (C.L).

A) Normalized standard protein curve					
C.L	Concentration (ng/µL)	I.V. (µL)	Cg (ng)	Matrix (ng)	BSA (ng)
1	5.0	2	0	0	10
2	7.5	2	5	0	10
3	10.0	2	10	0	10
4	15.0	2	20	0	10
5	20.0	2	30	0	10
6	30.0	2	50	0	10

In Table 3.10 the ratio between the area of each proteotypic peptide (numbered as 1, 2, 3, 4 and 5 according to the Table 3.2) and the area of the corresponding BSA peptide (n-area) were reported for each replicate analysis together with the correspondent n-area average, standard deviation, and RSD % at each calibration level. The RSD % of the peptide 1 ranged from 5.08 to 20.06 %; the RSD % of peptide 2 from 1.06 to 14.61 %, peptide 3 from 6.71to 25.87 %, peptide 4 from 6.54 to 35.44 % and finally peptide 5 from 4.96 to 18.65 %.
Table 3.11 reported the Replicate area ratios (n-Replicate area) calculated by averaging the area ratios of all proteotypic peptide (numbered as 1, 2, 3, 4, and 5 according to Table 3.2) within replicate 1, 2 or 3 at each calibration level; the three n-Replicate areas at each calibration level were averaged to obtain the calibration level area ratios (n-C.L. area). In Table 3.11 the standard deviation (sd) and the RSD % of the calibration level area (n-C.L. area) were reported too. The normalized standard protein curve was obtained by plotting the six calibration level area (n-C.L. areas) vs. the ratios between nanograms of purified γ-conglutin and BSA loaded on chip at each calibration level. The RSD % of the calibration level area (n-C.L. area) ranged from 7.66 to 16.05 %.

Table3.10: Values for the preparation of normalized standard peptide curves. Normalized standard peptide curves are obtained by plotting the normalized peptide area averages (n-average) of each proteotypic peptide (numbered as 1, 2, 3, 4, and 5 according to Table 3.2) at each calibration level (C.L) vs. the ratios between the amount of γ-conglutin (Cγ/BSA ng/ng) and of BSA loaded on HPLC-Chip column. Peptide n-area averages (n-average) are obtained by averaging normalized peptide areas (n-area) of each replicate within each calibration level. The standard deviation (sd) and the RSD % of peptide area averages are reported too.

C.L.	Cγ/BSA	peptide 1				peptide 2				peptide3				peptide 4				peptide 5			
	(ng/ng)	n-area	n-average	sd	RSD%	n-area	n-average	sd	RSD%	n-area	n-average	sd	RSD%	n-area	n-average	sd	RSD%	n-area	n-average	sd	RSD%
1	0	0.0000	0.0000	0.0000	0.00	0.0000	0.0000	0.0000	0.00	0.0000	0.0000	0.0000	0.00	0.0000	0.0000	0.0000	0.00	0.0000	0.0000	0.0000	0.00
		0.1028				0.5175				0.3622				0.7076				1.1334			
2	0.5	0.1447	0.1305	0.0239	18.35	0.6071	0.5349	0.0653	12.20	0.4428	0.4109	0.0429	10.43	0.7957	0.7400	0.0484	6.54	1.2441	1.2811	0.1693	13.22
		0.1439				0.4801				0.4278				0.7169				1.4659			
		0.1879				0.8837				0.8944				0.6103				2.6050			
3	1	0.2052	0.2137	0.0310	14.49	0.8883	0.8913	0.0094	1.06	1.1202	1.0962	0.1908	17.41	1.2213	1.1018	0.1785	16.20	2.5240	2.6836	0.2102	7.83
		0.2481				0.9018				1.2738				1.4737				2.9217			
		0.4861				1.5464				2.9705				2.0763				4.9216			
4	2	0.5144	0.4848	0.0304	6.26	1.4313	1.4962	0.0589	3.94	3.2461	3.0224	0.2028	6.71	2.6525	2.4488	0.3230	13.19	6.9342	5.7223	1.0675	18.65
		0.4538				1.5108				2.8506				2.6175				5.3109			
		0.5959				2.0294				3.9749				1.6245				7.5555			
5	3	0.7842	0.6398	0.1283	20.06	2.5192	2.1595	0.3155	14.61	6.0257	4.6400	1.2003	25.87	2.7203	2.8274	1.0019	35.44	9.6497	8.2856	1.1823	14.27
		0.5391				1.9297				3.9195				4.1373				7.6518			
		0.8649				2.8462				7.0667				6.3594				14.6982			
6	5	0.9293	0.8792	0.0447	5.08	2.8944	2.8137	0.1010	3.59	6.6864	6.6374	0.4558	6.87	5.9262	5.2256	1.6033	30.68	14.4531	14.1758	0.7033	4.96
		0.8435				2.7005				6.1591				3.3913				13.3762			

Table 3.11: Values for preparing normalized standard protein curve. Normalized standard protein curve is obtained by plotting normalized calibration level areas (n-C.L. areas) vs. the ratio between the amount of γ-conglutin and of BSA (Cγ/BSA ng/ng) loaded on HPLC-Chip column. n-C.L. areas are calculated by averaging the three normalized Replicate areas (n-Replicate area) within each calibration level (C.L.). n-Replicate areas are obtained by averaging the peptide areas ratios (n-area) of all the five proteotypic peptides (numbered as 1, 2, 3, 4, and 5 according to Table 3.2) within each replicate (Rep. N°). Standard deviation (sd) and RSD % of n-C.L. area are reported too.

C.L.	Cγ/BSA (ng/ng)	Rep. N°	peptide n-area					n-Replicate area	n-C.L. area	sd	RSD%
			1	2	3	4	5				
1	0	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.00
		1	0.1028	0.5175	0.3622	0.7076	1.1334	0.5647			
2	0.5	2	0.1447	0.6071	0.4428	0.7957	1.2441	0.6469	0.6195	0.0475	7.66
		3	0.1439	0.4801	0.4278	0.7169	1.4659	0.6469			
		1	0.1879	0.8837	0.8944	0.6103	2.6050	1.0363			
3	1	2	0.2052	0.8883	1.1202	1.2213	2.5240	1.1918	1.1973	0.1638	13.68
		3	0.2481	0.9018	1.2738	1.4737	2.9217	1.3638			
		1	0.4861	1.5464	2.9705	2.0763	4.9216	2.4002			
4	2	2	0.5144	1.4313	3.2461	2.6525	6.9342	2.9557	2.6349	0.2876	10.92
		3	0.4538	1.5108	2.8506	2.6175	5.3109	2.5487			
		1	0.5959	2.0294	3.9749	1.6245	7.5555	3.1560			
5	3	2	0.7842	2.5192	6.0257	2.7203	9.6497	4.3398	3.7105	0.5954	16.05
		3	0.5391	1.9297	3.9195	4.1373	7.6518	3.6355			
		1	0.8649	2.8462	7.0667	6.3594	14.6982	6.3671			
6	5	2	0.9293	2.8944	6.6864	5.9262	14.4531	6.1779	5.9463	0.5727	9.63
		3	0.8435	2.7005	6.1591	3.3913	13.3762	5.2941			

Samples for the normalized in-matrix protein curve were obtained by digesting increasing amounts of the purified γ-conglutin in the presence both of a constant amount of the protein extract and of the internal standard BSA: constant volumes of protein extract were spiked both with increasing volumes of the BSA standard solution and increasing volumes of the dialyzed solution of the purified γ-conglutin obtaining tryptic digest with the final concentrations reported in Table 3.12. All samples were analyzed in three replicates using the implemented HPLC-Chip-MRM method by injecting 2 μl of each sample.
In Table 3.13 the normalized area (n-area) of each proteotypic peptide (numbered as 1, 2, 3, 4 and 5 according to the Table 3.2) for each replicate were reported together with the correspondent normalized area average (n-average), standard deviation (sd), and RSD % at each calibration level. The RSD % of the peptide 1 n-area in presence of the matrix ranged from 1.16 to 9.81 %; the RSD % of peptide 2 from 1.83 to 6.63 %, peptide 3 from 3.92 to 21.27 %, peptide 4 from 1.47 to 16.70 % and finally peptide 5 from 4.23 to 27.12 %.
Table 3.14 reported the normalized Replicate areas (n-Replicate area) calculated by averaging the normalized area (n-area) of all proteotypic peptides (numbered as 1, 2, 3, 4, and 5 according to Table 3.2) within replicate 1, 2 or 3 at each calibration level; the three n-Replicate areas were then averaged to obtain the normalized calibration level areas (n-C.L.area). In Table 3.14 the standard deviation (sd) and the RSD % of the n-C.L. area were reported too. The normalized in-matrix protein curve was obtained by plotting the six normalized-calibration areas vs. the ratio between the amount of purified γ-conglutin and of BSA (ng/ng) loaded on chip at each calibration level (i. e. 0, 0.5, 0.1, 0.20, 0.3 and 0.5 ng) and in presence of a constant amount of matrix (i.e. 200 ng, see Table 3.6). The RSD % of the normalizedcalibration level area ranged from 5.09 to 13.57 %

Table 3.12: Description of the normalized in-matrix protein curve samples: concentrations of the tryptic digests (ng/µL), injection volumes (I.V.), amounts of purified γ-conglutin (ng) and matrix (ng) loaded on the HPLC-Chip column at each calibration level (C.L).

B) Normalized in-matrix protein curve*					
C.L	Concentration (ng/µL)	I.V. (µL)	C-γ (ng)	Matrix (ng)	BSA (ng)
1	105.0	2	0	200	10
2	107.5	2	5	200	10
3	110.0	2	10	200	10
4	115.0	2	20	200	10
5	120.0	2	30	200	10
6	130.0	2	50	200	10

*In addition to the spiked γ-conglutin, these samples contained an unknown amount of endogenous γ-conglutin inside the matrix.

Table 3.13: Values for the preparation of normalized in-matrix peptide curves. Normalized in-matrix peptide curves are obtained by plotting the normalized peptide area averages (n-average) of each proteotypic peptide (numbered as 1, 2, 3, 4, and 5 according to Table 3.2) at each calibration level (C.L) vs. the ratios between the amount of γ-conglutin and BSA (Cγ/BSA ng/ng) loaded on HPLC-Chip column in presence of 200 ng of matrix. Normalized peptide area averages (n-average) are obtained by averaging normalized peptide areas (n-area) of each replicate within each calibration level. The standard deviation (sd) and the RSD % of peptide area averages (n-average) are reported too.

C.L.	Cγ/BSA (ng/ng)	peptide 1				peptide 2				peptide3				peptide 4				peptide 5			
		n-area	n-average	sd	RSD%	n-area	n-average	sd	RSD%	n-area	n-average	sd	RSD%	n-area	n-average	sd	RSD%	n-area	n-average	sd	RSD%
		0.2025				0.7150				0.8396				1.0726				2.8503			
1	0	0.1996	0.1960	0.0089	4.54	0.6573	0.6841	0.0291	4.25	1.0206	0.8842	0.1204	13.62	1.1389	1.0891	0.0440	4.04	3.2122	2.7285	0.5547	20.33
		0.1859				0.6801				0.7925				1.0557				2.1230			
		0.3251				1.0321				1.6994				1.7432				5.2282			
2	0.5	0.3119	0.3043	0.0254	8.35	0.9740	0.9902	0.0366	3.69	1.6327	1.7585	0.1636	9.31	1.7284	1.6790	0.0987	5.88	4.9350	5.3393	0.4699	8.80
		0.2760				0.9645				1.9435				1.5653				5.8548			
		0.4227				1.2545				2.8278				2.1598				5.6766			
3	1	0.4145	0.4074	0.0197	4.84	1.2657	1.2582	0.0065	0.52	3.2562	3.0585	0.2161	7.07	2.7755	2.3641	0.3563	15.07	4.2031	5.4766	1.1862	21.66
		0.3852				1.2544				3.0914				2.1569				6.5501			
		0.5422				1.8020				3.4161				2.6722				10.2700			
4	2	0.6218	0.5983	0.0488	8.16	1.9297	1.9649	0.1829	9.31	3.5204	3.1596	0.5372	17.00	2.9714	2.8533	0.1593	5.58	8.3342	8.6996	1.4233	16.36
		0.6309				2.1628				2.5422				2.9163				7.4946			
		0.8695				2.7654				3.9415				4.1079				13.5046			
5	3	0.7849	0.8442	0.0516	6.11	2.4536	2.6895	0.2086	7.76	6.9568	5.6156	1.5350	27.33	5.5758	4.6831	0.7837	16.74	15.3657	13.1024	2.4889	19.00
		0.8782				2.8496				5.9486				4.3655				10.4368			
		1.2961				3.6191				7.7052				6.1308				15.4941			
6	5	1.1862	1.2366	0.0555	4.49	3.5649	3.4703	0.2124	6.12	10.0902	9.2242	1.3198	14.31	6.9422	6.7676	0.5699	8.42	19.8295	19.3035	3.5755	18.52
		1.2276				3.2271				9.8772				7.2297				22.5868			

Table 3.14: Values for preparing normalized in-matrix protein curve. Normalized in-matrix protein curve is obtained by plotting normalized calibration level areas (n-C.L. areas) vs. the ratio between the amount of γ-conglutin and BSA (Cγ/BSA ng/ng) loaded on HPLC-Chip column in presence of 200 ng of matrix. n-C.L. areas are calculated by averaging the three normalized Replicate areas (n-Replicate area) within each calibration level (C.L.). n-Replicate areas are obtained by averaging the peptide n-areas of all the five proteotypic peptides (numbered as 1, 2, 3, 4, and 5 according to Table 3.2) within each replicate (Rep. N°). Standard deviation (sd) and RSD % of n-C.L. area are reported too.

C.L.	Cγ (ng)	Rep. N°	peptide n-area					n-Replicate area	n-C.L. area	sd	RSD%
			1	2	3	4	5				
		1	0.2025	0.7150	0.8396	1.0726	2.8503	1.1360			
1	0	2	0.1996	0.6573	1.0206	1.1389	3.2122	1.2457	1.1164	0.1402	12.56
		3	0.1859	0.6801	0.7925	1.0557	2.1230	0.9674			
		1	0.3251	1.0321	1.6994	1.7432	5.2282	2.0056			
2	0.5	2	0.3119	0.9740	1.6327	1.7284	4.9350	1.9164	2.0143	0.1025	5.09
		3	0.2760	0.9645	1.9435	1.5653	5.8548	2.1208			
		1	0.4227	1.2545	2.8278	2.1598	5.6766	2.4683			
3	1	2	0.4145	1.2657	3.2562	2.7755	4.2031	2.3830	2.5129	0.1571	6.25
		3	0.3852	1.2544	3.0914	2.1569	6.5501	2.6876			
		1	0.5422	1.8020	3.4161	2.6722	10.2700	3.7405			
4	2	2	0.6218	1.9297	3.5204	2.9714	8.3342	3.4755	3.4551	0.2961	8.57
		3	0.6309	2.1628	2.5422	2.9163	7.4946	3.1494			
		1	0.8695	2.7654	3.9415	4.1079	13.5046	5.0378			
5	3	2	0.7849	2.4536	6.9568	5.5758	15.3657	6.2273	5.3870	0.7313	13.57
		3	0.8782	2.8496	5.9486	4.3655	10.4368	4.8957			
		1	1.2961	3.6191	7.7052	6.1308	15.4941	6.8490			
6	5	2	1.1862	3.5649	10.0902	6.9422	19.8295	8.3226	8.0004	1.0289	12.86
		3	1.2276	3.2271	9.8772	7.2297	22.5868	8.8297			

Figure 3.5 reported normalized standard curves (I) and normalized in-matrix curves (II) for each proteotypic peptides (peptide 1 in plot A, peptide 2 in plot B, peptide 3 in plot C, peptide 4 in plot D, peptide 5 in plot E) of γ-conglutin. Figure 3.6 reported the normalized standard protein curve (I in plot F) and the normalized in-matrix protein curve (II in plot F).
The normalized standard peptide curves (AI, BI, CI, DI, and EI of Figure 3.3) were obtained by plotting the normalized area average of peptides vs. the ratio between the nanograms of the purified γ-conglutin and BSA loaded on chip at each calibration level (Table 3.10). The normalized standard protein curve (Figure 3.6 plot FI) was obtained by plotting the normalized calibration level areas vs the ratio between the nanograms of the purified γ-conglutin and of the BSA loaded on chip at each calibration level (Table 3.11). It showed a very good linearity in the experimental range, with an excellent regression coefficient (R^2=0.998) and an intercept with the y-axis very close to 0. The statistical features of the normalized standard protein curve (Figure 3.6 plot FI) appeared to be as satisfactory as those of the standard protein curve (Figure 4 plot FI).
The normalized in-matrix peptide curves (AII, BII, CII, DII, and EII of Figure 3.5) were obtained by plotting the normalized area averages (n-area) of peptides vs. the ratio between the nanograms of the purified γ-conglutin and the BSA nanograms loaded on chip at each calibration level and in presence of a constant amount of matrix (i.e. 200 ng) (Table 3.13). The normalized in matrix protein curve were prepared by plotting the normalized calibration level areas (n-C.L. area) vs. the ratio between the nanograms of the purified γ-conglutin and the BSA nanograms. It was characterized by a better linear behavior than the corresponding in matrix protein curve, prepared in absence of the internal standard BSA, having a regression coefficient equal to 0.995. The normalization procedure appears to be responsible for the observed increased linearity.

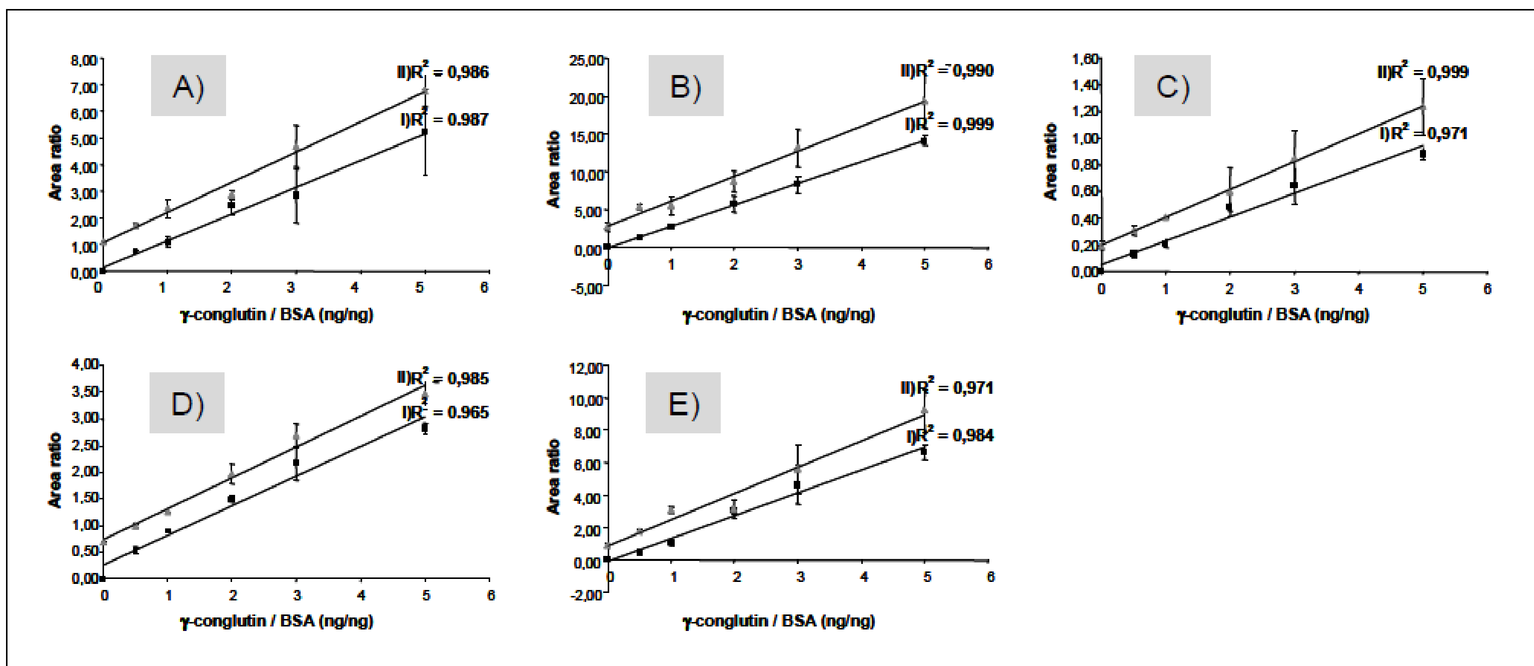


Figure 3.5: Normalized standard peptides curves (I) and normalized in-matrix peptide curves (II) of peptide 1 (plot A), 2 (plot B), 3 (plot C), 4 (plot D) and 5 (plot E).

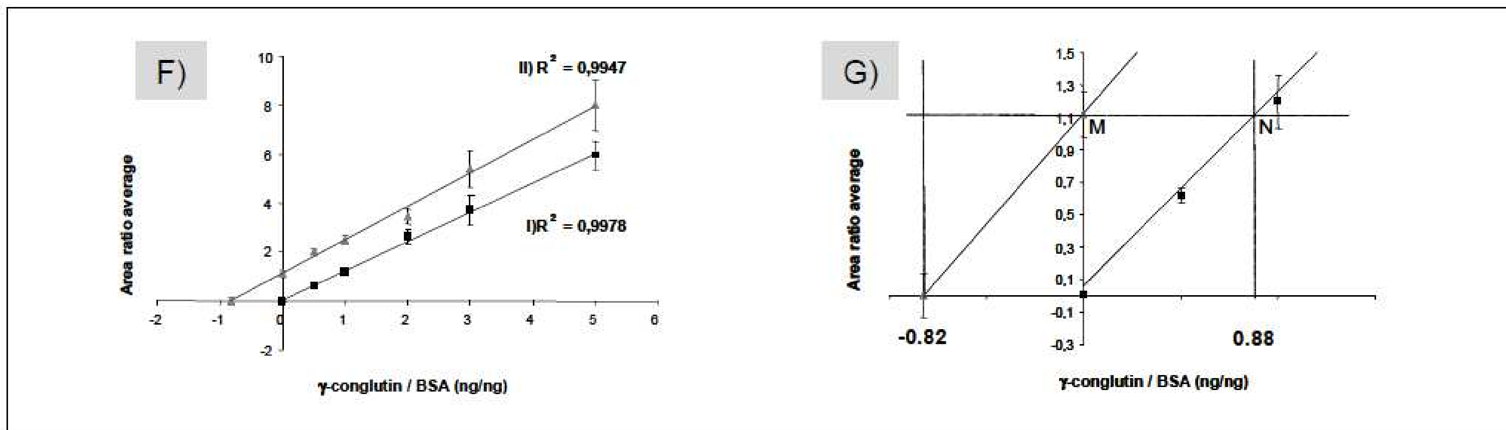


Figure 3.6: Plot F) Normalized standard protein curve (I) and normalized in-matrix protein curve (II); Plot G) Evaluation of the absolute quantification of γ-conglutin obtained from the normalized standard protein curve and normalized in-matrix protein curve.

The comparison of each normalized in-matrix peptide curve vs. the corresponding normalized standard peptide curve (Figure 3.5, plots AI-II, BI-II, CI-II, DI-II, and EI-II) permitted to affirm that the normalization procedure was able to compensate the matrix effect on each proteotypic γ- conglutin peptide. Being all the normalized peptide curve couples parallel, also the normalized in-matrix protein curve was parallel to the normalized standard protein curve (see curves FI-II in Figure 6 in comparison with Figure 3.4).

On the basis of the parallelism of the normalized in-matrix protein curve with the normalized standard protein curve (Figure 3.6, plot FI-II), it was possible to directly quantify the absolute amount of endogenous γ-conglutin in the matrix (200 ng) using the normalized standard protein curve as an external calibration curve. The absolute amount of γ-conglutin in 200 ng of matrix was firstly calculated by using the normalized in-matrix protein curve (Figure3.6, curve FII). The intercept of this curve with the X-axis gave the value of -0.82 ± 0.102 for the ratio between γ-conglutin and BSA (expressed as nanograms/nanograms) (Figure 3.6, plot G).

Since a constant amount of BSA equal to10 ng had been injected, the resulting absolute amount of endogenous γ-conglutin in 200 ng of matrix was 8.2±1.02 ng. Moreover, the projection on the x-axis of point N (interception between the normalized standard protein curve and a line parallel to the X-axis passing for point M, i.e. the interception of the normalized in-matrix protein curve with the y-axis) gave a value of 0.88 for the same ratio, corresponding to 8.8 ng of γ-conglutin. This value was very close to that of 8.2 ng obtained from the normalized in-matrix protein curve. The agreement of the two calculated values confirmed the reliability of the direct quantification of the γ-conglutin using the normalized standard protein curve as the external calibration curve. To unquestionably confirm this conclusion, a new sample at the calibration level 1 of the normalized in matrix protein curve (Table 3.12) was prepared and injected in triplicate in order to really verify the reliability of the normalized standard protein curve as external calibration curve, which was now used to quantify the endogenous γ-conglutin. The obtained value was equal to 8.6 ± 0.91 ng of γ- conglutin in 200 ng of matrix. On the contrary, the estimation of the γ-conglutin content using the two non-normalized protein curves (i.e. the standard protein curve and the standard protein curve in Figure 3.4, plot FI-II) was not reliable as showed by the plot G in Figure 3.4.

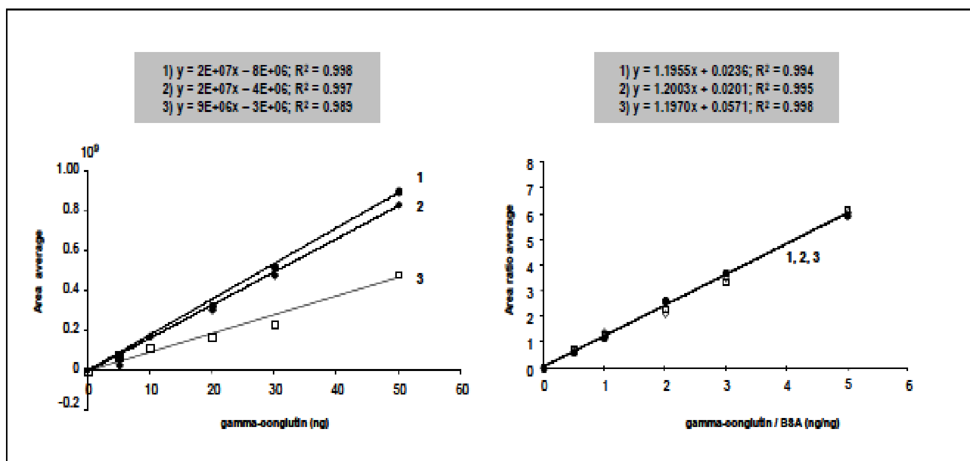


Figure 3.7: Plot A) Three replicate standard protein curves; Plot B) Three replicate normalized protein curves.

In order to verify the inter-day reproducibility of the method, the normalized protein curves were prepared three times in different days (time 0, after 5 days, and after 30 days). The three non-normalized standard protein curves (Figure 3.7 A) had quite different slopes, whereas the three normalized standard protein curves were perfectly super imposable (Figure 3.7 B), demonstrating a very satisfactory inter-day reproducibility.

3.5 Conclusion of study 2

Study 2 may be considered as the development and the optimization of label-free absolute quantitative method for the validation of the differential analysis of γ-conglutin in study 1. In fact, it is based on the proteotypic peptides of this protein which are selectively monitored in Multiple Reaction Monitoring (MRM) mode. The use of the HPLC-Chip-Ion Trap system, working in MRM enabled the detection of five γ-conglutin proteotypic peptides starting from 5 ng of purified γ-conglutin loaded on chip, correspondent to a concentration of the tryptic digest equal to 2.5 ng/ul. At this concentration, all five proteotypic peptides were quantifiable exceeding the common threshold of LOQ equal to 10.

The normalized standard protein curve prepared by detecting in Multiple Reaction Monitoring mode the five proteotypic peptides of γ-conglutin and normalizing their area respect to the area of two internal standard BSA peptides, was shown to be a reliable and robust external calibration curve. According to my knowledge, this is the first time in which the standard addition approach is applied to the development of a label-free absolute quantification in proteomics, by using a strategy based on proteotypic peptides. The main limitation of the presented method regards the availability of a purified sample of the target protein: not all proteins are commercially available and, sometimes, their quantitative purification may be difficult or very time consuming. This relatively cheap approach seems to be more suitable to nutrition and food science than to biomedicine: food matrices may be less complex and the concentrations of the target proteins may be higher than those of biomarkers.

3.6 References

Blonder J, Veenstra TD, 2007, Computational prediction of proteotypic peptides. Nat Biotechnol 25:125-131.

Ito S, Tsukada K, 2001, Matrix effect and correction by standard addition in quantitative liquid chromatographic-mass spectrometric analysis of diarrhetic shellfish poisoning toxins. J Chrom A, 943:39-46.

Maggi F et al., 2009, *Melittis melissophyllum* L. subsp. melissophyllum (Lamiaceae) from central Italy: A new source of a mushroom-like flavour. Food Chem 113:216-221.

Magni C., et al., 2007, Combined 2D electrophoretic approaches for the study of white lupin mature seed storage protome. Phytochem 68:997-1007.

Pascoe R et al., 2001, Reduction in Matrix-Related Signal Suppression Effects in Electrospray Ionization Mass Spectrometry Using On-Line Two-Dimensional Liquid Chromatography. Anal Chem 73:6014-6023.

Sanders WS et al., 2007, Prediction of peptides observable by mass spectrometry applied at the experimental set level. BMC Bioinformatics 1:8.

4. Conclusion

Functional foods and beverages have the appearance of normal foods, but contain specific components whose activity on at least one measurable risk factor has been scientifically demonstrated. In some cases the nutraceutical properties depend on proteins or peptides. Literature reports different examples of food proteins characterized by different biological activities. In order to evaluate the nutraceutical value of a functional food, it is certainly essential to develop methods able to quantify each bioactive component.

.Lupin is a functional ingredient characterized by a high nutritional value and good technological flexibility that only recently has attracted the interest of research: experimental and clinical investigations have indicated that lupin proteins may be useful for controlling hypercholesterolemia [Sirtori et al., 2004, Spielmann et al., 2007] and hyperglycemia [Magni et al., 2004, Lee et al., 2006].

Traditionally, target proteins are quantified by immunoenzymatic o electrophoretic methods, which, however, they have some drawbacks: a) their success relies on the time consuming production and validation of specific antibodies; b) some immunoassays are not sensitive enough and may respond to a family of proteins rather than to single target protein; c) the lack of specific antigens may cause cross-reactivity problems and false positive results [Murthy et al., 1998]; and d) the possibility of adapting immunological methods to multiplex analysis remains limited [Rifai et al., 2006]. A good alternative to the classical immunoenzymatic methods may be provide by shotgun proteomics [Mamone et al., 2009], an analytical technique based on liquid chromatography coupled with mass mass spectrometry. In particular, the high-performance liquid chromatography (HPLC) coupled with Ion Trap mass spectrometry via an electrospray source has become a powerful technique to develop high specific, sensitive and accurate quantitative methods because of the high resolving power of the chromatography and the selectivity and sensitivity of the mass spectrometry.

In food analysis stable isotope labeling (SIL) techniques appear to be too expensive, whereas proteomic tools based on stable isotope label-free (SIF) techniques may find some important applications owing to their simple experimental workflows and capability of comparing an unlimited number of samples.

In study 1 the proposed shotgun-proteomics analysis based on HPLC-Chip-MS/MS allowed a complete characterization of lupin seed storage proteins, since minor proteins, such as gamma-conglutin and delta-conglutin, were easily identified together with major proteins with satisfactory percentage coverages without any previous fractionation of the Total Protein Extract (TPE).

The homology mode search has permitted to identify single aminoacid substitution in vicilin and delta conglutin sequences. In particular, the MS/MS sequencing of single aminoacid substituted peptides, starting from the vicilin-like protein and the beta-conglutin precursor, acquiring important knowledge of the heterogeneous nature of vicilin isoforms. Moreover the label free differential analysis was able to profile different expression of vicilins and gamma conglutin among cultivars of white lupin. Two different algorithm for the relative quantification of gamma conglutin and vicilins the ”normalized protein mean peptide spectral intensity” (N-Mean) and the “normalized protein average of common reproducible peptides” (N-ACRP), were developed.

N-MEAN appears to be the most suitable parameter for profiling the differential expression of vicilin class, since being the mean peak intensity of all peptide precursor ions identified, it permits to take into consideration all vicilin isoforms identified in the chromatograms.
On the contrary, the normalized parameter N-ACRP appears to be the most suitable parameter for profiling the differential expression of gamma-conglutin. It was shown to have a good reliability and a precision close to those obtained using strategies relying on chemical or metabolic labeling.
The study 2 may be considered as a progress of the study 1.
According to the theory of proteotypic peptides exposed by Anderson [Anderson et al., 2006, Mallick et al., 2007] in the study 2 the label free method for the absolute quantification of gamma conglutin was developed by selectively monitored five proteotypic peptides in multiple reaction monitoring (MRM) mode.The originality of this study was the "translation" of the Standard addition strategy from the analytical area to the label free quantitative proteomics.
The main analytical problem of the label-free quantitative approaches in mass spectrometry concerned the well-know matrix effect which affected the detection of the analytes co-eluting with several matrix components. The BSA normalization procedure assured an optimal reliability of the label-free method since it was able to minimize the matrix-effect. The compensation of the matrix effect in the γ-conglutin quantification has been demonstrated by the direct comparison between the normalized in-matrix protein curve and the normalized standard protein curve, both prepared on the standard addition principle, which resulted to be perfectly parallel. The final output of the optimized method was a reliable and robust external calibration curve, the normalized standard protein curve, which was obtained by averaging the normalized standard curves of five target γ-conglutin peptides, detected in a very selective MRM mode. The five target peptides assured a good sequence coverage of the γ-conglutin; the quantification of a target protein considering more than one peptides is more reliable because using a single peptide standard, any sequence variation or post-translational modification in the unique marker peptide will dismiss the quantification of the corresponding protein.
Study 1 and 2 are based on the peptide on the peak intensity label-free parameter. This may be considered either an advantage or a disadvantage in quantitative approach: peptide area parameter is more suitable than the others to highlight small change in protein amount among sample but simultaneously it is more affected by the effect exerted to the matrix on the peptide detection (matrix effect). The use of an internal standard protein to normalized quantitative parameter and the use of a reproducible chromatographic system such as HPLC-Chip appear to be very useful to increase the reliability of quantitative analysis based on peak area.
Old and coworker [Old et al., 2005] performed a comparison of the spectral counting versus the peak intensity procedure for the protein quantification. They demonstrated that peak intensity measurements displayed more accurate estimates of protein ratio. Working in data dependent acquisition mode the sensitivity of peak detection was limited to those peptides selected in MS/MS.
The medium complexity of samples in study 1 and 2 allowed working in monodimensional liquid chromatography. The high sensitivity of methods has been reached by working in nano-flow and nano-electrospay by using microfluidic HPLC-Chip system. It assured a high reproducibility of chromatographic separation resulting in a high reproducibility of the peptide areas and the retention time both working in MS/MS and in MRM mode. For this reason a previous alignment of multiple LC-MS/MS runs was not necessary. Considering the medium complexity of the samples and the high performance of the HPLC-Chip system the acquisition of MS/MS spectra using a classical 3D ion trap with a limited capacity and low scan speed has not been a limitation in study 1 and 2.
In order to obtained highly quantitative results statistical analysis was performed on the data , thi requires replicate injections per sample. All samples of study 1 and 2 were injected in three

technical replicates in order to evaluate the reproducibility of the data and obtained reliable evaluation on linearity of considered parameters.
The use of an exogenous internal standard protein used to spike samples prior to the tryptic digestion enabled to obtained quantitative data. In study 1 the reliability of the two algorithms N-MEAN for vicilins and N-ACRP for gamma conglutin has been demonstrated.
The normalization of peak area with respect to standard co eluting peptides enabled to reach the parallelism between the normalized standard protein curve in study 2. The normalization procedure using an exogennoous internal standard minimized the matrix effect that was easily estimated by comparing the standard protein curve and in-matrix protein curve in study 2.
In conclusion, this thesis has demonstrated that it is possible to develop label-free methods for the quantification of nutraceutical protein or peptides in food matrices.

4.1 References

Anderson L., Hunter C. L., 2006, Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol. Cell. Proteomics 5:573-588.

Lee YP et al., 2006, Lupin-enriched bread increases satiety and reduces energy intake acutely. Am J Clin Nutr 84:975-980.

Magni C et al., 2004, Conglutin gamma, a lupin seed protein, binds insulin in vitro and reduces plasma glucose levels of hyperglycemic rats. J Nutr Biochem 15:646-650.

Marthy JN et al., 1998, Tacrolimus metabolite cross-reactivity in different tacrolimus assay. Clin Biochem 31:613-617.

Mallick P et al., 2007, Computational prediction of proteotypic peptides for quantitative proteomics. Nat Biotechnol 25:125-131.

Mamone G, 2009, Analysis of food proteins and peptides by mass spectrometry-based techniques. J Chromatogr A. 1216:7130-7142.

Old et al., 2005, Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cel Proteomics 4:1487-1502.

Rifai et al., 2006, Proptein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotechnol 24:971-983.

Sirtori CR et al., 2004, Proteins of white lupin seed, a naturally isoflavone-poor legume, reduce colesterolemia in rats and increase LDL receptor activity in HepG2 cells. J Nutr 134:18-23.

Spielmann et al., 2007, Dietary lupin protein lowers triglyceride concentrations in liver and plasma in rats by reducing hepatic gene expression of sterol regulatory element-binding protein-1c. Ann Nutr Metab, 51:387-392.

Appendix 1: Abstract

Proteomic techniques offer a new approach for the characterization of food ingredients. Food quality is, in fact, dependent on the presence of bioactive proteins which could have either beneficial or negative effects on human health. For example, quantitative proteomics based on mass spectrometry has been used for detecting and quantifying allergenic proteins or bioactive compounds in tiny amounts. In order to profile the differential protein expression in different samples, two main approaches are reported in literature: stable isotope labeling (SIL) techniques and stable isotope label-free (SIF) techniques. In food analysis, in particular, SIL techniques appear to be too expensive, whereas proteomic tools based on SIF techniques may find some important applications owing to their simple experimental workflows and capability of comparing an unlimited number of samples.

In the study 1, an internal standard label-free method based on ion intensity for the simultaneous identification and relative quantification of target storage proteins in total protein extracts (TPEs) of the seeds of *Lupinus albus* (white lupin) was developed. The use of an innovative microfluidic system, the HPLC-Chip, coupled with a classical Ion Trap mass spectrometer has enabled a complete qualitative characterization of all seed storage proteins in a single analysis of the TPE tryptic digest.

The differential analyses of γ-conglutin, a mature protein, and of the vicilins, a complex protein class, in four lupin cultivars were performed optimizing two suitable bioinformatics parameters, the "normalized protein average of common reproducible peptides" (N-ACRP) and the "normalized protein mean peptide spectral intensity" (N-MEAN), respectively.

It is important to underline that a relative approach does not enable the absolute quantification of the target protein in the sample and, consequently, lacks to give a real evaluation of the potential bioactivity of the food. The quantitative approaches AQUA and QConCAT, requiring the chemical synthesis of all isotope-labeled peptides, appear not suitable to their application in the field of food chemistry. Moreover, in food analysis the isotope labelling techniques appear to be too expensive. In order to achieve a real absolute quantification of the lupin γ-conglutin, in study 2, a very selective method was develop and applied to TPE. The Multiple Reaction Monitoring (MRM) label-free absolute quantitative method, based on the "standard addition" strategy was developed with the target to absolutely quantify the lupin target protein, i.e. γ-conglutin, in the flour of white lupin.

The four main features of the method are the following: a) the chromatographic separation was performed on a very efficient HPLC-Chip system coupled with a ion trap mass spectrometer; b) five proteotypic peptides of γ-conglutin were selected and analyzed with a Multiple Reaction Monitoring (MRM) method; c) the absolute quantification was obtained by the standard addition approach by purifying γ-conglutin from lupin seed; d) the matrix effect was overcome by the addition of an exogenous protein.

Appendix 2: Riassunto

Le tecniche proteomiche offrono un nuovo approccio per la caratterizzazione degli ingredienti alimentari. La qualità di un alimento, infatti, dipende dalla presenza di proteine bioattive che potrebbero avere sia effetti negativi, sia effetti positivi sulla salute. Per esempio la proteomica quantitativa viene utilizzata per la determinazione e la quantificazione di proteine allergeniche o composti bioattivi in tracce. Al fine di delineare il profilo proteico due principali approcci vengono riportati in letteratura: le tecniche "stable isotope labeling" (SIL) e le tecniche "stable isotope label-free" (SIF). Nell'analisi degli alimenti le tecniche SIL appaiono troppe costose, al contrario le tecniche SIF trovano applicazione per la semplicità del loro schema di lavoro sperimentale e la capacità di confrontare potenzialmente un illimitato numero di campioni.

Nel primo studio è stato sviluppato un metodo label-free basato sull'intensità degli ioni per la simultanea identificazione e quantificazione relativa di proteine di riserva nell'estratto proteico totale (TPE) del seme di lupino bianco (*Lupinus albus*). L'utilizzo di un innovativo sistema microfluidico HPLC-Chip accoppiato ad una classica trappola ionica, assicura una completa caratterizzazione qualitativa di tutte le proteine di riserva del seme in una singola analisi di un digerito triptico di TPE. E' stata effettuata l'analisi differenziale della conglutina gamma, una proteina matura, e delle viciline, una classe proteica eterogenea, in quattro cultivar di lupino e sono stati ottimizzati due parametri bioinformatica, N-MEAN e N-ACRP.

E' importante sottolineare che un approccio relativo non è in grado di dare una quantificazione assoluta di una proteina target in un campione e, di conseguenza, questo potrebbe portare ad un errata quantificazione della potenziale bioattività di un alimento. Nella proteomica quantitativa assoluta, gli approcci quantitativi AQUA e QConCAT necessitano della sintesi di tutti i peptidi marcati; l'applicazione di queste tecniche appare quindi non idonea a essere utilizzata nel campo della chimica degli alimenti. Inoltre nell'analisi di un alimento le tecniche che prevedono l'utilizzo di isotopi marcati appaiono troppo costose. Al fine di ottenere la quantificazione assoluta della conglutina gamma, nello studio 2 è stato sviluppato un metodo label free in modalità Multiple Reaction Method (MRM), estremamente selettivo. Tale metodo basato sull'approccio delle addizioni standard, è stato applicato al TPE con lo scopo di effettuare una quantificazione assoluta della conglutina gamma nella farina di lupino. Le principali caratteristiche del metodo sono: a) un'efficiente separazione cromatografica grazie ad un efficace sistema HPLC-Chip accoppiato ad uno spettrometro di massa a trappola ionica; b) la scelta di cinque peptidi proteotipici; c) la purificazione della conglutina gamma a partire dal seme di lupino e conseguente quantificazione assoluta ottenuta mediante l'approccio delle addizioni standard; d) l'aggiunta di uno standard interno per compensare l'effetto matrice.

Appendix 3: Papers

3.1 Paper 1

Francesca Brambilla, Donatella Resta, Ilena Isak, Marco Zanotti, Anna Arnoldi, 2009, A label-free internal standard method for the differential analysis of bioactive lupin proteins using nano HPLC-Chip coupled with Ion Trap mass spectrometry. Proteomics, 9:272-286.

3.2 Paper 2

Elena Sirtori, Donatella Resta, Francesca Brambilla, Christian Zacherl, Anna Arnoldi, 2010, The effects of various processing conditions on a protein isolate from *Lupinus angustifolius*. Food Chemistry, 120:496-504.

3.3 Paper 3

Elena Sirtori, Donatella Resta, Anna Arnoldi, Huub F.J. Savelkoul, Harry J. Wichers, 2010, Cross-reactivity between peanut and lupin proteins, Food Chemistry, in press, accepted for publication on 12nd November 2010.

3.4 Paper 4

Donatella Resta, Francesca Brambilla, Marco Zanotti, Anna Arnoldi, 2010, HPLC-Chip-Multiple Reaction Monitoring (MRM) method for the label-free absolute quantification of γ-conglutin in lupin: proteotypic peptides and standard addition method. J Proteome Research, submitted.

Appendix 4: Abstract of posters

4.1 Poster 1

Brambilla F., Resta D., Isak I., Boschin G., Zanotti M., Arnoldi A. "HPLC-Chip-Ion Trap label-free method method for the differential analysis of the major bioactive lupin proteins." 26th Informal Meeting on Mass Spectrometry, 4-8 maggio 2008, Fiera di Primiero. (Poster)

4.2 Poster 2

Resta D., Brambilla F., Boschin G., Zanotti M., Arnoldi A. "Development of quantitative mass spectrometric multiple reaction monitoring assay for major lupin allergens." 26th Informal Meeting on Mass Spectrometry,4-8 maggio 2008. Fiera di Primiero. (Poster)

4.3 Poster 3

Arnoldi A., Resta D., Brambilla F. "Label-free differential analysisi of gamma-conglutin in different cultivars of *Lupinus albus* using nano HPLC-Chip coupled with Ion Trap Mass Spectrometry." 10th International Symposium On Immunological, Chemical and Clinical Problems Of Food Allergy, 26-29 maggio 2008, Parma. (Poster)

4.4 Poster 4

Resta D., Brambilla F., Zanotti M., Arnoldi A. "Preliminary approaches to the development of a label-free absolute quantification of gamma-conglutin in complex protein mixtures." Italian annual association 3th Annual National Conference, 11-14 giugno 2008, Selva di Fasano. (Poster)

4.1 Abstract poster 1

HPLC-CHIP-ION TRAP "LABEL FREE" METHOD FOR THE DIFFERENTIAL ANALYSIS OF THE MAJOR BIOACTIVE LUPIN PROTEIN.

Francesca Brambilla[1], *Donatella Resta*[1], *Giovanna Boschin*[1], *Marco Zanotti*[2] *and Anna Arnoldi*[1]

[1]Institute of Endocrinology, Laboratory of Food Chemistry and Mass Spectrometry, University of Milan, Via Balzaretti 9, 20133 Milan, Italy
[2]Agilent Technologies Italia SpA, Via Gobetti 2/C, 20063 Cernusco sul Naviglio, Milan, Italy

Recent and very promising applications of proteomics have been provided in the field of food quality analysis to monitor changes in specific food protein components, such as bioactive or allergenic proteins. Mass spectrometry-based quantitative proteomics has become an important tool for food science. In order to profile the differential protein expression in different samples by mass spectrometry, two main approaches are reported in literature: stable isotope labeling techniques (SIL) and stable isotope label free techniques (SIF). However SIL techniques remains the core technology used in mass spectrometry-based quantification of plasma biomarker of disease, risk and therapeutic response, increasing efforts have been directed to the label-free approaches in the field of food allergens. Label-free techniques allow to compare the relative protein abundances in an unlimited number of samples by acquiring independent data and by comparing any data set to any other data set. Moreover, label-free approach is attractive for its simplicity as well as cost effectiveness. The aim of this work was to develop an innovative label-free method based on shotgun proteomics for the simultaneous identification and relative quantification of lupin bioactive protein in total protein extracts of different cultivars of *Lupinus albus*, a grain legume which is gaining in the interest of food industry. The introduction of an exogenous internal standard protein at a constant level in the protein mixtures subjected to enzymatic digestion seems to be an interesting resolution both for evaluating matrix effect and for normalizing quantitative parameter of target proteins, in addition to the use of the very sensitive and reproducible chromatographic system named HPLC-Chip.

4.2 Abstract poster 2

DEVELOPMENT OF QUANTITATIVE MASS SPECTROMETRIC MULTIPLE REACTION MONITORING ASSAY FOR MAJOR LUPIN ALLERGENES

Donatella Resta[1], *Francesca Brambilla*[1], *Giovanna Boschin*[1], *Marco Zanotti*[2] *and Anna Arnoldi*[1]

[1]Institute of Endocrinology, Laboratory of Food Chemistry and Mass Spectrometry, University of Milan, Via Balzaretti 9, 20133 Milan, Italy
[2]Agilent Technologies Italia SpA, Via Gobetti 2/C, 20063 Cernusco sul Naviglio, Milan, Italy

Gamma-conglutin is a mature protein composed by a heavy and a light chain linked by disulfide bonds. It is a bioactive protein with a putative hypoglycemic activity. It also seems to be the major allergen. This hypothesis is supported by some peculiar molecular properties such as thermal stability and resistance to proteolysis that could be considered important features frequently shared by food allergens. This prompted the European Commission to include this seed in the list of food allergens whose declaration on food label is compulsory. Nowadays, the quantification of food allergens is generally based on immunoassays. These methodologies, however, have some limitations, such as cross-reactivity with other food proteins and false-positive results.
The aim of our work is the development of a reliable quantitative HPLC-Chip-Ion Trap method for the absolute quantification of the gamma conglutin that could be apply to food and food ingredient quality evaluation.
The peptide detection sensitivity using Multiple Reaction Monitoring (MRM) approach is expected to be greater than that achieved in a full scans MS data dependent approach. Therefore, a MRM method for the quantification of lupin gamma-conglutin in complex protein mixtures is optimized. Specific tryptic peptides are selected as stoichiometric representative of target protein and quantified against a spiked internal standard (bovine serum albumin, BSA) to provide absolute quantification of the protein concentration.

4.3 Abstract poster 3

LABEL-FREE DIFFERENTIAL ANALYSIS OF GAMMA CONGLUTIN IN DIFFERENT CULTIVARS OF *LUPINUS ALBUS* USING NANO HPLC-CHIP COUPLED WITH ION TRAP MASS SPECTROMETRY.

Anna Arnoldi, Donatella Resta and Francesca Brambilla.

Institute of Endocrinology, Laboratory of Food Chemistry and Mass Spectrometry, University of Milan, Via Balzaretti 9, 20133 Milan, Italy.

There are a few literature indications that some individuals are allergic to lupin proteins. This prompted the European Commission to include this seed in the list of food allergens whose declaration on food labels is compulsory.
Considering that a few specific investigations have demonstrated that gamma-conglutin may be one of the major lupin allergens, the detection of this protein is a main analytical issue. The hypothesis is supported by some peculiar structural properties of this protein, such as thermal stability and resistance to proteolysis that are important features frequently shared by food allergens.
In order to profile the differential protein expression in different samples by mass spectrometry, two main approaches are reported in literature: stable isotope labeling techniques (SIL) and stable isotope label free techniques (SIF). Although SIL techniques remains the core technology used in mass spectrometry-based quantification of plasma biomarker of disease, risk and therapeutic response, in the field of food allergens increasing efforts have been directed to label-free approaches. Label-free techniques allow to compare the relative protein abundances in an unlimited number of samples by acquiring independent data and by comparing any data set to any other data set. Moreover, label-free approach is attractive for its simplicity as well as cost effectiveness.
The aim of this work was to develop an innovative ion current-based label-free method for the simultaneous identification and relative quantification of gamma-conglutin in total protein extracts (TPE) of different cultivars of *Lupinus albus* (*cv.* Adam, Ares, Lucky, Multitalia). Intensity-based label-free quantitation is not generally accepted as reliable without the use of an internal standard for the normalization of the considered quantitative parameters. Therefore the TPEs (cv. Adam, Ares, Lucky, Multitalia) were spiked with an internal standard protein (bovine serum albumin, BSA) at a constant level prior to the enzymatic digestion. This was useful for minimizing the matrix effect and for normalizing the quantitative parameters Pr-ACRP optimized for gamma conglutin. The spiked protein mixtures were analyzed by HPLC-Chip-MS/MS without any preliminary separation to preserve the actual composition of starting materials.

4.4 Abstract poster 4

PRELIMINARY APPROACHES FOR THE DEVELOPMENT OF A "LABEL-FREE" ABSOLUTE QUANTIFICATION OF GAMMA CONGLUTIN IN COMPLEX PROTEIN MIXTURE.

Donatella Resta[1], Francesca Brambilla[1], Marco Zanotti[2] and Anna Arnoldi[1].

[1]Institute of Endocrinology, Laboratory of Food Chemistry and Mass Spectrometry, University of Milan, Via Balzaretti 9, 20133 Milan, Italy
[2]Agilent Technologies Italia SpA, Via Gobetti 2/C, 20063 Cernusco sul Naviglio, Milan, Italy

Quantitative proteomics based on mass spectrometry has been used in food quality control for detecting and quantifying proteins in tiny amounts which could have either negative or beneficial effects on human health: allergens and bioactive proteins respectively.
In literature two main approaches are reported in order to profile the relative abundance of target proteins in different samples: stable isotope labeling (SIL) techniques and stable isotope label-free (SIF) techniques. SIL tecniques, in spite of their potency, have some limitations such as the use of expensive labeled reagents; therefore they are certainly not applicable in the area of food analysis.
The actual increasing interest for lupin proteins is based both on their nutritional and technological characteristics, that permit to use them as ingredients in the formulation of a large range of different food products, and on their potential nutraceutical properties. In particular gamma-conglutin, a mature seed storage protein composed by a heavy and a light chain linked by disulfide bonds, seems to be the hypoglicemic component and recent studies have demonstrated that it may be the major allergen in lupin seed.
The aim of this work was the development of a HPLC-Chip mass spectrometric Multiple Reaction Monitoring (MRM) assay for the quantification of *Lupinus albus* gamma conglutin in complex protein mixture. Specific tryptic peptides are selected as stoichiometric representative of the target protein and quantify against a spiked internal standard (bovine serum albumin, BSA) to provide absolute quantitation of protein concentration.